Applications du transport optimal à des problèmes de limites de champ moyen

François Bolley

Ecole Normale Supérieure de Lyon
Unité de Mathématiques Pures et Appliquées

Thèse

présentée en vue de l'obtention du grade de

Docteur de l'Ecole Normale Supérieure de Lyon

Discipline : Mathématiques

par

François Bolley

Applications du transport optimal à des problèmes de limites de champ moyen

---

Directeur de thèse : M. Cédric Villani

Soutenue publiquement le 5 décembre 2005 après avis de

M. Michel Ledoux,	*Professeur - Université de Toulouse III*
M. Félix Otto,	*Professeur - Universität Bonn*

devant le jury composé de

M. Yann Brenier,	*Directeur de recherche CNRS - Université de Nice*
M. José Antonio Carrillo,	*Professeur - Universitat Autònoma de Barcelona*
M. Patrick Cattiaux,	*Professeur - Université de Paris X*
M. Michel Ledoux,	*Professeur - Université de Toulouse III*
M. Cédric Villani,	*Professeur - Ecole Normale Supérieure de Lyon*

Remerciements

Je souhaite tout d'abord exprimer ma profonde gratitude à Cédric Villani pour avoir encadré cette thèse et pour la confiance qu'il m'a accordée. Il a su me faire découvrir et approfondir tant de sujets passionnants ; j'ai beaucoup appris auprès de lui. De plus sa bonne humeur et son enthousiasme communicatifs ont été de grandes sources de motivation tout au long de ces trois années de doctorat.

Michel Ledoux et Félix Otto ont bien voulu être rapporteurs de cette thèse ; c'est un honneur pour moi et je tiens à leur témoigner ici toute ma reconnaissance. Qu'ils soient également remerciés pour l'excellent accueil qu'ils m'ont réservé au sein de leurs équipes à Toulouse et à Bonn, leur disponibilité et l'intérêt qu'ils ont montré pour mon travail.

Yann Brenier, José Antonio Carrillo et Patrick Cattiaux me font l'honneur de participer au jury et je les en remercie. Leurs travaux sont de grandes sources d'inspiration et je tiens aussi à leur exprimer mes sincères remerciements pour les discussions fructueuses que nous avons eues à plusieurs reprises.

Pendant la préparation de cette thèse j'ai aussi eu la possibilité de collaborer avec Gordon Blower, Arnaud Guillin et Grégoire Loeper. Je les remercie vivement de ces collaborations, et espère avoir de nouveau le plaisir de travailler avec eux.

Je tiens également à exprimer toute ma gratitude à Benoît Perthame pour sa grande disponibilité à l'Ens : il a dirigé mes premiers pas dans le domaine et a toujours montré de l'intérêt pour l'avancement de mon travail.

C'est aussi pour moi un plaisir de remercier Luis Caffarelli, Irene Gamba et Takis Souganidis de m'avoir chaleureusement accueilli dans leur équipe lors d'un séjour à Austin au printemps 2003.

Je n'oublie pas bien entendu tous les membres de l'Umpa qui, par leur disponibilité et leur ouverture, ont su créer une atmosphère de travail et de relations en tous points excellente. Un merci particulier à Edouard et Matthieu pour leur précieuse aide technique, Mylène pour l'animation du bureau, et Ana, Aurélien, Lolo, Thom, Tefa et Tomasz pour les activités sportives.

Table des matières

Introduction

L'état d'un système physique constitué d'un très grand nombre de particules soumises à diverses actions peut être étudié à plusieurs niveaux, c'est-à-dire suivant des échelles (d'espace, de temps par exemple) différentes.

On peut tout d'abord en étudier un *modèle microscopique*, c'est-à-dire un modèle mathématique dans lequel l'inconnue est l'état de *chaque* particule. Ces modèles peuvent être donnés sous la forme d'équations d'évolution : suivant les systèmes considérés, et suivant les paramètres intervenant dans la description de l'état des particules - position, vitesse, ... - les équations, écrites dans un espace des phases très grand, pourront par exemple prendre la forme des équations de Newton pour des particules classiques en interaction, d'un flot gradient pour des systèmes dissipatifs, ... De tels modèles prennent donc la forme de systèmes d'équations nombreuses et couplées, ce qui peut rendre leur étude difficile. C'est pourquoi on préfère souvent remplacer cette description microscopique de chaque particule dans un espace des phases très grand par une description *macroscopique* dans un espace des phases réduit.

Ainsi, dans une description macroscopique *cinétique*, l'état du système est à chaque instant représenté par une densité de présence dans l'espace des phases $\mathbb{R}^d \times \mathbb{R}^d$ des positions x et vitesses v. Cette densité peut obéir à une équation aux dérivées partielles d'évolution posée sur $[0, +\infty[\times \mathbb{R}^d \times \mathbb{R}^d$, dont la forme traduit le type d'interaction entre les particules. Si seules les particules suffisamment proches l'une de l'autre interagissent, la densité vérifie une équation de Boltzmann ; si au contraire chaque particule ressent l'influence de toutes les autres particules, alors l'équation satisfaite par la densité est dite *de champ moyen* : c'est le cas par exemple de l'équation de Vlasov de la forme

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + F(t,x) \cdot \nabla_v f \;=\; 0,$$

dont l'inconnue $f(t,x,v)$ est la densité de présence au temps t dans l'espace $\mathbb{R}^{2d}$ des positions x et des vitesses v ; c'est aussi le cas, pour des systèmes homogènes en position dont l'état est donné par une densité de présence dans l'espace des vitesses uniquement, de certaines équations de McKean-Vlasov de la forme

$$\frac{\partial f}{\partial t} = \Delta_v f + \nabla_v \cdot (f \nabla_v V) + \nabla_v \cdot (f(\nabla_v W * f)),$$

dont l'inconnue $f(t,v)$ est la densité de présence au temps t dans l'espace $\mathbb{R}^d$ des vitesses v.

Dans une description macroscopique *hydrodynamique*, l'état du système est représenté par des fonctions sur l'espace des positions uniquement donnant à chaque instant t la densité, la vitesse et la température dans l'espace des positions x, obtenues comme des moments en vitesse v de la densité cinétique de présence $f(t,x,v)$. Par exemple, les équations d'Euler

$$\begin{cases} \dfrac{\partial u}{\partial t} + (u \cdot \nabla)\, u &= -\nabla p \\ \nabla \cdot u &= 0 \end{cases}$$

régissent l'évolution du champ de vitesse $u(t,x)$ d'un fluide non visqueux en régime incompressible.

Dans le travail présenté ici, nous nous intéressons essentiellement à des modèles macroscopiques cinétiques. Une fois le modèle obtenu, sous forme d'une équation aux dérivées partielles sur la densité de présence, il peut être intéressant de chercher à en approcher les solutions. Plusieurs types de méthodes sont envisageables, dont la précision peut dépendre de la forme de l'équation considérée. Nous étudions ici des *méthodes d'approximation particulaires* pour des équations de champ moyen ; elles consistent en l'introduction d'un grand nombre N de particules fictives, évoluant selon un système d'équations différentielles couplées, ordinaires ou stochastiques, posé dans un espace très grand, mais dans un sens plus simple à résoudre que l'équation macroscopique.

Ainsi dans le cas de l'équation de Vlasov, ces N particules fictives, repérées par leur position et vitesse $(X_t^{i,N}, V_t^{i,N})$ pour $1 \leq i \leq N$ dans l'espace des phases $\mathbb{R}^{2d}$, évoluent suivant les équations de Newton régissant la dynamique des particules physiques du modèle microscopique. Il ne s'agit cependant en rien d'un retour vers ce modèle : en effet, si le système physique original est constitué d'un nombre de particules de l'ordre de 10^{25} par exemple, l'espoir de la méthode est de pouvoir approcher de manière satisfaisante l'état du système donné par la densité macroscopique de présence, et même l'état du système physique des 10^{25} particules, par l'introduction d'un nombre N limité de particules fictives $(X_t^{i,N}, V_t^{i,N})$, de l'ordre de 10^6 par exemple. L'état du système de particules introduites par la méthode d'approximation est décrite par les observables $\displaystyle\frac{1}{N}\sum_{i=1}^{N} \varphi(X_t^{i,N}, V_t^{i,N})$ pour des fonctions test φ, et peut donc être décrit par une mesure de probabilité, dite *mesure empirique*, définie sur l'espace des phases $\mathbb{R}^{2d}$ par

$$\hat{\mu}_t^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{(X_t^{i,N}, V_t^{i,N})}.$$

Dans le cas de l'équation de McKean-Vlasov, les particules fictives introduites dans l'approximation particulaire sont repérées par leur seule vitesse dans $\mathbb{R}^d$ et évoluent suivant des équations différentielles stochastiques conduites par un mouvement brownien rendant compte du terme de diffusion donné, au niveau macroscopique de l'équation, par le laplacien. On définit de même une mesure empirique dans l'espace des positions.

L'étude de la précision de la méthode d'approximation particulaire se ramène donc à la comparaison de deux mesures de probabilité, dépendant du temps, dont l'une (la solution de l'équation macroscopique) est déterministe, alors que l'autre (la mesure empirique) peut être déterministe ou aléatoire suivant le modèle. La convergence de la mesure empirique vers la densité macroscopique, pour N tendant vers l'infini, est appelée *limite de champ moyen* pour les équations macroscopiques considérées. Il s'agit d'en donner des formulations précises, avec si possible l'obtention d'estimations explicites et l'étude de leur dépendance en temps.

Son analyse repose en particulier sur une bonne connaissance des propriétés des solutions des équations d'évolution (des particules d'une part, de la densité de présence d'autre part) intervenant dans le problème, mais aussi sur l'introduction d'outils analytiques traduisant les distances entre ces mesures : deux exemples fondamentaux en sont la distance de Wasserstein (liée à la théorie du transport optimal de mesures) et la notion d'entropie relative ou d'information de Kullback. Nous rappelons en particulier, dans la partie 0 de cette introduction, quelques notations et quelques propriétés des distances de Wasserstein dont nous ferons usage dans la suite.

L'introduction se compose ensuite de trois parties présentant les différents thèmes et résultats des travaux composant cette thèse. Dans la partie I nous abordons l'étude de limites de champ moyen pour les équations de Vlasov et d'Euler incompressibles dans le plan : dans ces modèles le problème de limite se ramène à une question de stabilité d'équations aux dérivées partielles, qui peut être résolue par des propriétés de type contraction en distances de Wasserstein. Nous présentons aussi un résultat analogue pour des lois de conservation scalaires. Dans la partie suivante nous nous intéressons à des inégalités de concentration, ou de déviation, de certaines mesures de probabilité, et à leurs liens avec les inégalités de transport (qui lient les distances de Wasserstein et l'entropie) et les inégalités de Sobolev logarithmiques ; en particulier nous donnons de telles inégalités pour des lois jointes de variables dépendantes. Enfin, dans la partie III, nous considérons une approximation d'équations de McKean-Vlasov par des systèmes de particules évoluant suivant des équations différentielles stochastiques ; nous donnons une estimation quantitative de la précision de la méthode à l'aide des inégalités de concentration introduites dans la partie précédente.

0 - Distances de Wasserstein

Ces distances, qui évaluent l'écart entre deux mesures de probabilité, sont liées au problème de transport optimal, formulé originalement par G. Monge de la manière suivante. Soient, dans le plan $\mathbb{R}^2$, deux distributions μ_0 et μ_1 de même masse d'un matériau donné. Sachant que le coût du transport d'une unité de masse d'un point x à un point y du plan est la distance euclidienne $|x-y|$ dans $\mathbb{R}^2$, comment doit-on transporter la première distribution sur la deuxième de manière à minimiser le coût de transport donné par

$$\int_{\mathbb{R}^2} |t(x) - x| \, \mu_0(x) \, dx$$

si tout point x de la distribution μ_0 est envoyé au point $t(x)$ de la distribution μ_1 ?

Dans le cadre plus abstrait d'un espace polonais X, d'un coût de transport $c(x,y)$ pour aller de x à y et de deux mesures boréliennes de probabilité μ_0 et μ_1 sur X, il s'agit de trouver, parmi toutes les applications t de X dans X, dites de transport, envoyant μ_0 sur μ_1 dans le sens où μ_1 est la mesure image de μ_0 par t notée $t\sharp\mu_0$, c'est-à-dire dans le sens où

$$\mu_0[t^{-1}(B)] = \mu_1[B]$$

pour tout borélien B de X, celles qui minimisent la quantité

$$\int_X c(x, t(x))\, d\mu_0(x).$$

Ce problème n'a pas toujours de solution : par exemple, sur $\mathbb{R}$ muni de la distance usuelle, si μ_0 est la masse de Dirac en 0 et μ_1 est la moyenne des masses de Dirac en -1 et $+1$, il n'existe même pas d'application t envoyant μ_0 sur μ_1, puisque le point $x = 0$ doit être envoyé à la fois sur $y = -1$ et $y = +1$.

On est alors amené à la formulation plus faible suivante du problème due à L. Kantorovich : minimiser la quantité

$$\iint_{X\times X} c(x,y)\, d\pi(x,y)$$

parmi tous les éléments π, appelés plans de transport, de l'ensemble $\Pi(\mu_0, \mu_1)$ des mesures boréliennes de probabilité sur l'espace produit $X \times X$ de marginales μ_0 en la première variable et μ_1 en la deuxième. On autorise ainsi à répartir la masse $d\mu_0(x)$ initialement en x en divers points y selon la masse $d\pi(x,y)$. Cet ensemble $\Pi(\mu_0, \mu_1)$ est toujours non vide car il contient le produit tensoriel $\mu_0 \otimes \mu_1$. Ce nouveau problème est bien une formulation faible du problème initial puisque d'une part toute application de transport t induit un plan de transport π défini par $\pi = (Id \times t)\sharp\mu_0$ où Id est l'application identité sur X, et d'autre part les deux infima sont égaux sous certaines conditions sur c et μ_0. Il admet toujours une solution dès que c est une fonction semi-continue inférieurement sur $X \times X$, et sous certaines hypothèses supplémentaires sur X, c et μ_0, ce plan de transport optimal est unique et s'écrit en fait sous la forme $(Id \times t)\sharp\mu_0$ pour une certaine application t.

Ces résultats et plus généralement l'étude de ce problème de transport et ses liens avec de nombreux domaines, tels que l'analyse de certaines équations aux dérivées partielles, des inégalités fonctionnelles ou géométriques, ou la théorie des probabilités, sont exposés dans [1], [93], [94] ou [111] par exemple. Ici nous ne rappelons que quelques notations et résultats dont nous nous servirons à plusieurs reprises dans la suite.

Dans le cas où le coût c est donné par $c(x,y) = d(x,y)^p$ où d est une distance sur X semi-continue inférieurement sur $X \times X$ et p est un nombre strictement positif, on pose

$$W_p(\mu_0, \mu_1) = \inf_{\pi\in\Pi(\mu_0,\mu_1)} \Big(\iint_{X\times X} d(x,y)^p\, d\pi(x,y) \Big)^{\min(1,1/p)}$$

pour deux mesures boréliennes μ_0 et μ_1 de probabilité sur X, cette quantité étant éventuellement infinie. Elle définit une distance sur l'espace $\mathcal{P}_p(X)$ des mesures boréliennes de probabilité μ sur X telles que le moment $\displaystyle\int_X d(x,x_0)^p\, d\mu(x)$ soit fini pour un (et donc tout)

$x_0 \in X$, notée W_p et appelée *distance de Wasserstein* ou de Monge-Kantorovich d'ordre p associée à d.

Si par exemple d est la distance triviale donnée par $d(x,y) = \mathbf{1}_{x \neq y}$, la distance W_1 d'ordre 1 correspondante vérifie la relation

$$W_1(\mu_0, \mu_1) = \frac{1}{2} \|\mu_0 - \mu_1\|_{TV}$$

où $\|\mu_0 - \mu_1\|_{TV}$ est la norme de variation totale de la mesure $\mu_0 - \mu_1$.

De façon générale la distance W_1 satisfait la formulation duale de Kantorovich-Rubinstein

$$W_1(\mu_0, \mu_1) = \sup_{[\varphi]_{lip} \leq 1} \Big\{ \int_X \varphi(x)\, d\mu_0(x) - \int_X \varphi(x)\, d\mu_1(x) \Big\} \tag{1}$$

(cf. [50] ou [111, Theorem 1.13] par exemple), où $[\varphi]_{lip}$ désigne la semi-norme de Lipschitz définie par

$$[\varphi]_{lip} = \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x,y)}.$$

Dans le cas où $X = \mathbb{R}$ muni de la distance usuelle et $p \geq 1$, la distance $W_p(\mu_0, \mu_1)$ entre deux mesures μ_0 et μ_1 peut être exprimée simplement à l'aide des pseudo-inverses F_i^{-1} des fonctions de répartition F_i des μ_i, définies sur $[0,1]$ par

$$F_i^{-1}(t) = \inf \big\{ x \in \mathbb{R}; F_i(x) > t \big\}$$

où

$$F_i(x) = \int_{-\infty}^{x} d\mu_i(y) = \mu_i \big[ ]-\infty, x] \big]$$

pour $x \in \mathbb{R}$ et $i = 0, 1$. On a en effet la relation

$$W_p(\mu_0, \mu_1) = \Big(\int_0^1 |F_0^{-1}(t) - F_1^{-1}(t)|^p dt \Big)^{1/p} \tag{2}$$

qui, pour $p = 1$, s'écrit aussi

$$W_1(\mu_0, \mu_1) = \int_{-\infty}^{+\infty} |F_0(x) - F_1(x)| dx \tag{3}$$

par le théorème de Fubini.

Au niveau topologique, si d est une distance métrisant X pour laquelle il est complet et si W_p est définie à partir de d, nous verrons en particulier dans l'annexe de ce mémoire que la convergence d'une suite $(\mu_n)_n$ de $\mathcal{P}_p(X)$ vers une mesure de probabilité μ pour W_p est, sous une condition de moments, équivalente à la convergence faible (étroite) de $(\mu_n)_n$ vers μ. Nous y montrerons également que l'espace $(\mathcal{P}_p(X), W_p)$ est aussi un espace polonais, complet pour la distance W_p (cf. **Theorem A.2**).

I - Propriétés de type contraction de certaines équations aux dérivées partielles

Dans cette partie I nous abordons l'étude de limites macroscopiques par l'approximation de deux équations, d'une part l'équation de Vlasov et d'autre part l'équation d'Euler incompressible dans le plan en formulation vorticité, par des systèmes de particules déterministes. Notant alors que la mesure empirique de ces particules est elle-même une solution (faible) de l'équation de Vlasov ou d'une version régularisée de l'équation d'Euler, le problème de la limite de champ moyen se ramène à un problème de stabilité de solutions de ces équations : sachant qu'une donnée initiale g_0 (ici la mesure empirique du système de particules en leurs positions initiales) est proche d'une autre donnée initiale f_0 fixée (ici la densité initiale de présence dans l'espace des phases pour l'équation de Vlasov ou la vorticité initiale du fluide pour l'équation d'Euler), est-on sûr qu'en chaque instant ultérieur $t > 0$ la valeur g_t à l'instant t de la solution correspondante g de donnée initiale g_0 soit encore proche de la valeur f_t de la solution f de donnée initiale f_0 ?

Un tel problème de stabilité peut, dans certains cas favorables, être énoncé de manière précise et résolu de façon simple. Ainsi nous verrons dans le paragraphe I.1 de cette introduction pour l'équation de Vlasov avec potentiel lipschitzien et dans le paragraphe I.2 pour l'équation d'Euler régularisée que pour $p = 1$ et 2 il existe une constante positive L_p telle que, pour toutes solutions f et g de données initiales f_0 et g_0 respectives ayant certains moments finis, on ait la relation

$$W_p(f_t, g_t) \leq e^{L_p t}\, W_p(f_0, g_0), \qquad t \geq 0$$

où W_p est la distance de Wasserstein d'ordre p. En particulier, par exemple dans le cas de l'équation de Vlasov, étant donnés T et $\varepsilon > 0$, la mesure empirique du système de particules au temps t sera éloigné d'au plus ε (en distance W_p) d'une solution donnée f_t, uniformément sur $[0, T]$, si initialement la mesure empirique est à distance au plus $\varepsilon\, e^{-L_p T}$ de f_0.

Ce type de résultat quantitatif de stabilité a été étudié et montré pour plusieurs classes d'équations, dans des sens plus ou moins forts. Ainsi par exemple H. Tanaka [108] a obtenu le résultat de contraction

$$W_2(f_t, g_t) \leq W_2(f_0, g_0), \qquad t \geq 0$$

pour des solutions, de centre de masse et d'énergie cinétique fixés, de l'équation de Boltzmann homogène pour des molécules maxwelliennes (cf. aussi [19] et [111]) ; cela lui permet alors d'établir la convergence de ces solutions vers des distributions gaussiennes. De même J. A. Carrillo, M. P. Gualdani et G. Toscani [35], [38] obtiennent cette même propriété pour les distances de Wasserstein d'ordre $2p$ avec p entier, entre solutions de l'équation des milieux poreux en dimension 1; le résultat s'étend alors à la distance W_∞ qui mesure l'écart entre les supports de mesures.

Dans le paragraphe I.3 nous considérerons une loi de conservation scalaire, visqueuse ou non, et montrerons qu'en chaque instant $t \geq 0$ les dérivées spatiales $\partial u_t/\partial x$ et $\partial \tilde{u}_t/\partial x$ de deux solutions u et $\tilde{u}$ croissant en variable d'espace de 0 à 1, qui sont donc des mesures de probabilité, vérifient l'inégalité

$$W_p\Big(\frac{\partial u_t}{\partial x}, \frac{\partial \tilde{u}_t}{\partial x}\Big) \leq W_p\Big(\frac{\partial u_0}{\partial x}, \frac{\partial \tilde{u}_0}{\partial x}\Big), \qquad t \geq 0 \tag{4}$$

pour tout $p \geq 1$. Comme, pour $p = 1$, la distance W_1 entre deux mesures de probabilité sur $\mathbb{R}$ s'écrit comme la norme L^1 de la différence de leurs fonctions de répartition, nous retrouverons en particulier la propriété de contraction

$$\|u_t - \tilde{u}_t\|_{L^1(\mathbb{R})} \leq \|u_0 - \tilde{u}_0\|_{L^1(\mathbb{R})}, \qquad t \geq 0$$

démontrée par S. Kružkov et qui est fondamentale dans la théorie de ces équations.

Notons que des résultats plus fins, de la forme

$$W_2(f_t, g_t) \leq \Phi(t)\, W_2(f_0, g_0), \qquad t \geq 0$$

avec $\Phi(t)$ tendant vers 0 quand t tend vers l'infini, ont été obtenus pour certaines classes de solutions d'équations très particulières. Il ne s'agit alors plus tant d'un résultat de stabilité que d'un résultat de comportement asymptotique : si en effet on peut prendre pour g_0 une solution stationnaire ou un profil remarquable de l'équation, cela signifie que f_t converge vers ce profil quand t tend vers l'infini, a priori en distance W_2, mais parfois pour des normes plus fortes grâce à d'autres arguments, avec un taux de convergence donné par la fonction Φ. La convergence sera dite exponentielle (resp. polynomiale) si $\Phi(t)$ est de la forme $C\, e^{-\lambda t}$ (resp. $C\, t^{-\lambda}$) avec $\lambda > 0$. Citons en particulier des résulats de convergence polynomiale obtenus par H. Li et G. Toscani [71] pour des équations dites de friction modélisant des milieux granulaires, ou par J. A. Carrillo, M. Di Francesco et G. Toscani [34] dans le cas où g est une solution autosimilaire de l'équation des milieux poreux (appelée profil de Barenblatt) et f une solution de même centre de masse. En dimension quelconque des résultats de convergence exponentielle ont été démontrés pour des équations de McKean-Vlasov par J. A. Carrillo, R. J. McCann et C. Villani dans [36, 37] sous des hypothèses de convexité sur les potentiels considérés (étendant ainsi des résultats connus pour l'équation de Fokker-Planck) ; nous reviendrons sur ces équations dans la dernière partie de cette introduction puisque nous en étudierons une approximation particulaire.

Remarquons que ces derniers résultats de contraction s'accompagnent souvent de conditions nécessaires sur les données initiales. Ainsi par exemple, si l'équation préserve le centre de masse, alors $\displaystyle\int_{\mathbb{R}^d} x\, df_0(x) - \int_{\mathbb{R}^d} x\, dg_0(x) = \int_{\mathbb{R}^d} x\, df_t(x) - \int_{\mathbb{R}^d} x\, dg_t(x)$, et la formulation de Kantorovich-Rubinstein assure que cette différence entre les centres de masses est bornée par $W_1(f_t, g_t)$; comme cette distance tend vers 0 quand t tend vers l'infini, il s'ensuit que nécessairement f_0 et g_0 ont même centre de masse. C'est ainsi qu'apparaissent des conditions imposées dans les travaux [34] (qui justement précise des résultats obtenus dans [35] sans ces conditions), [36] ou [108], par exemple sur les centres de masse et éventuellement l'énergie cinétique. Dans le cas de l'équation de Vlasov, nous verrons dans le paragraphe I.1 qu'une condition sur le centre de masse et la vitesse moyenne des distributions initiales permet d'affiner les résultats.

Les techniques utilisées pour obtenir de tels résultats de contraction sont de deux ordres.

Tout d'abord en dimension 1, les démonstrations font toutes appel à la formulation simple (2) des distances W_p à l'aide des fonctions de répartition, ou plutôt de leurs pseudo-inverses. Ainsi l'idée utilisée dans [34], [35], [38] et [71] est d'écrire (du moins formellement) l'équation vérifiée par le peudo-inverse F_t^{-1} de la fonction de répartition F_t de la solution considérée f_t ; contrôler la distance W_p d'ordre p entre deux solutions de l'équation initiale revient alors à contrôler la norme $L^p(]0,1[)$ de la différence entre les solutions d'une autre équation, parfois plus simple à étudier. Dans l'étude des lois de conservation scalaires (cf. paragraphe I.3) nous ne pouvons établir d'équation sur le pseudo-inverse que dans un cas très particulier où la méthode des caractéristiques s'applique, et pour lequel l'inégalité de contraction (4) est d'ailleurs une égalité. Nous obtenons alors le résultat dans le cas général en intégrant ce premier cas dans un schéma de discrétisation en temps.

Par contre en dimension supérieure, où on ne peut plus considérer les pseudo-inverses, les seuls résultats de ce type semblent avoir été démontrés en distances W_p d'ordre 1 et 2. En effet la distance W_1 s'avère d'utilisation facile grâce à la formulation de Kantorovich-Rubinstein (1). Cependant, si elle est pratique et permet souvent s'obtenir des résultats de manière simple, elle donne rarement des résultats optimaux. Pour affiner les résultats il convient alors d'utiliser la distance W_2.

Une structure différentielle adaptée sur l'espace $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ a en effet été formellement présentée dans les travaux précurseurs de F. Otto (cf. [87] en particulier), et permet d'interpréter certaines équations aux dérivées partielles comme un flot gradient sur la mesure solution : c'est le cas par exemple des équations de la chaleur, des milieux poreux, de Fokker-Planck et de McKean-Vlasov. Ce formalisme permet de deviner de manière heuristique des propriétés sur les solutions de telles équations. La construction précise se fait alors à l'aide d'une discrétisation en temps du flot gradient, qui elle ne nécessite pas de structure différentielle sur $\mathcal{P}_2(\mathbb{R}^d)$, mais uniquement la structure métrique donnée par W_2 (cf. aussi les travaux [2, 3], [33], [62], [64] et [86]).

Cette interprétation, associée à des techniques de dissipation d'entropie liées aux inégalités de Sobolev logarithmiques, a permis par exemple dans [36, 37] d'obtenir les résultats de convergence exponentielle pour des équations de McKean-Vlasov, de la forme

$$W_2(f_t, g_0) \leq e^{-\lambda t}\, W_2(f_0, g_0)$$

où g_0 est une solution stationnaire. Elle est également à la base de résultats de convergence de type polynomial pour l'équation de Burgers avec viscosité, de la forme

$$\|f_t - g_t\|_{L^1(\mathbb{R})} \leq C\, t^{-1/2}\, \|f_0 - g_0\|_{L^1(\mathbb{R})}$$

où g est une onde de diffusion particulière (cf. [45]).

Ce formalisme a aussi permis l'obtention de plusieurs résultats liés aux inégalités de Sobolev logarithmiques et de Talagrand que nous aborderons dans la partie II de cette introduction.

C'est enfin sur cette structure, à travers la théorie développée par L. Ambrosio, N. Gigli et G. Savaré dans [2, 3], que se baseront nos calculs menant aux résultats de stabilité en distance W_2 pour les équations de Vlasov et d'Euler que nous présenterons dans les paragraphes I.1 et I.2. Notons que, d'après [111, Section 8.3], ces équations doivent être vues comme des flots hamiltoniens et non des flots gradients dans cette structure : pour ces équations il ne s'agit alors pas d'établir une convergence vers un équilibre, mais seulement un résultat de stabilité du type

$$W_2(f_t, g_t) \le e^{L_2 t}\, W_2(f_0, g_0).$$

C'est ce que nous allons présenter maintenant, en commençant par l'équation de Vlasov.

I.1. Limite de champ moyen pour l'équation de Vlasov (cf. chapitre 1)

Dans ce paragraphe nous nous intéressons au problème de la limite de champ moyen pour l'équation de Vlasov, qui consiste en l'étude de la concordance éventuelle des descriptions d'un système physique réalisées d'une part à travers l'introduction d'une densité de présence évoluant selon l'équation de Vlasov, d'autre part à travers un grand nombre de particules évoluant selon les équations de Newton.

Etant donné un potentiel V sur $\mathbb{R}^d$, cette équation cinétique de Vlasov, dont l'inconnue est une probabilité de présence $f = f(t,x,v)$ dans l'espace des phases $\mathbb{R}^d \times \mathbb{R}^d$, dépendant du temps, s'écrit

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f - (\nabla_x V * \rho[f_t]) \cdot \nabla_v f = 0, \qquad t > 0, \qquad x, v \in \mathbb{R}^d \tag{5}$$

où, en notant de façon classique par $f_t : (x,v) \mapsto f(t,x,v)$ la densité de présence au temps t dans l'espace des phases, $\rho[f_t]$ est la densité de présence au temps t dans l'espace physique $\mathbb{R}^d$ des positions, c'est-à-dire la marginale de f_t donnée par

$$\rho[f_t](x) = \int_{\mathbb{R}^d} f_t(x,v)\, dv, \qquad x \in \mathbb{R}^d.$$

Ainsi l'évolution de la densité f_t en chaque point de l'espace des phases dépend de la valeur de la densité en tous les autres points à travers le terme de convolution (en x) $\nabla_x V * \rho[f_t]$: dans ce sens l'équation (5) est dite *de champ moyen.*

Un cas très intéressant est celui de l'équation de Vlasov-Poisson en dimension $d = 3$, utilisée en physique des plasmas, dans lequel le champ électrique dérive du potentiel $V(x) = (4\,\pi|x|)^{-1}$. Un autre cas intéressant est celui de l'équation correspondant au potentiel de gravitation $V(x) = -(4\,\pi|x|)^{-1}$, qui est utilisée dans la modélisation de la matière stellaire. Le problème de la limite de champ moyen dans ce cas de potentiels singuliers est encore ouvert, malgré les avancées récentes de M. Hauray et P.-E. Jabin [60].

Dans la suite nous nous restreindrons à des potentiels V lipschitziens, ou plus précisément de classe $\mathcal{C}^1$ à dérivées bornées, et tels que $\nabla_x V(0) = 0$. Nous appellerons solution de (5) avec donnée initiale f_0 dans $\mathcal{P}(\mathbb{R}^{2d})$ une fonction f continue de $[0, +\infty[$ dans $\mathcal{P}(\mathbb{R}^{2d})$ muni de la topologie faible (étroite), vérifiant l'équation au sens des distributions en (t, x, v) et valant f_0 en $t = 0$.

Dans ce cadre l'existence et l'unicité de solutions ont été établies dans [25] ou [104, Chapter 5] par exemple si la fonction $\nabla_x V$ est de plus bornée sur $\mathbb{R}^d$, l'unicité étant complétée d'un résultat de stabilité très fort sur lequel nous reviendrons par la suite.

Le modèle microscopique associé consiste en un système de N particules dont l'état est déterminé par les N couples $(X_t^{i,N}, V_t^{i,N})$ de $\mathbb{R}^{2d}$, évoluant suivant les équations de Newton

$$\left\{ \begin{array}{lcl} \dfrac{dX_t^{i,N}}{dt} & = & V_t^{i,N} \\ \dfrac{dV_t^{i,N}}{dt} & = & -\dfrac{1}{N} \displaystyle\sum_{j=1}^{N} \nabla_x V(X_t^{i,N} - X_t^{j,N}) \end{array} \right. \qquad 1 \leq i \leq N \tag{6}$$

à partir de données initiales déterministes $(X_0^{i,N}, V_0^{i,N})$. Des données initiales aléatoires pourraient également être considérées comme dans [25] ou [32]. Notons que l'existence et l'unicité des solutions de (6) sont assurées dès que $\nabla_x V$ est une fonction lipschitzienne.

L'état du système au temps t est donné par le N-uplet $((X_t^{1,N}, V_t^{1,N}), \dots, (X_t^{N,N}, V_t^{N,N}))$ de $(\mathbb{R}^{2d})^N$; cependant, quand N devient grand, la position de chaque particule dans l'espace des phases n'a que peu d'importance, et ne sont intéressantes que des quantités statistiques (observables) telles que la position physique moyenne $\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} X_t^{i,N}$, la vitesse moyenne $\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} V_t^{i,N}$ ou l'énergie cinétique $\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} |V_t^{i,N}|^2$. Or toutes ces quantités peuvent être obtenues en testant la valeur $\hat{\mu}_t^N$ à l'instant t de la *mesure empirique* $\hat{\mu}^N$ du système, définie par

$$\hat{\mu}_t^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{(X_t^{i,N}, V_t^{i,N})},$$

sur des fonctions particulières, telles que $(x, v) \mapsto x, v$ ou $|v|^2$ pour les observables précédentes. En d'autres termes, de la connaissance de cette mesure découle la connaissance de l'état du système. Cette mesure représente en fait au temps t une densité de particules dans l'espace des phases dans le sens où $\hat{\mu}_t^N(A)$ est la portion de particules dans la configuration $((X_t^{1,N}, V_t^{1,N}), \dots, (X_t^{N,N}, V_t^{N,N}))$ se trouvant dans l'ensemble A de l'espace des phases.

Notons enfin que la mesure empirique $\hat{\mu}_t^N$ est une mesure de probabilité sur $\mathbb{R}^{2d}$ quel que soit le nombre N de particules : il sera alors envisageable d'étudier la convergence de $\hat{\mu}_t^N$ dans $\mathcal{P}(\mathbb{R}^{2d})$ quand N tend vers l'infini.

Le problème de la limite macroscopique s'énonce alors ainsi : soient d'une part f_0 une mesure de probabilité sur $\mathbb{R}^{2d}$ évoluant en f_t selon l'équation de Vlasov, et d'autre part $(X_0^{i,N}, V_0^{i,N})$ pour $1 \leq i \leq N$, N points de l'espace des phases évoluant en $(X_t^{i,N}, V_t^{i,N})$ suivant les équations de Newton. Si la répartition des $(X_0^{i,N}, V_0^{i,N})$ approche f_0 dans le sens où leur mesure empirique $\hat{\mu}_0^N$ est proche de f_0, est-il alors vrai qu'en chaque instant t la solution f_t est encore bien approchée par la mesure empirique $\hat{\mu}_t^N$ des N couples $(X_t^{i,N}, V_t^{i,N})$?

Ainsi, si d est une distance sur $\mathcal{P}(\mathbb{R}^{2d})$, métrisant par exemple la topologie faible (étroite) sur cet espace, est-il vrai que la distance $d(\hat{\mu}_t^N, f_t)$ converge vers 0 quand N tend vers l'infini pour chaque instant $t \geq 0$ s'il en est ainsi en $t = 0$?

Comme dans le cadre macroscopique de l'équation de champ moyen (5), la force exercée sur chacune des N particules est à chaque instant la moyenne des forces exercées par toutes les autres particules du système, et non seulement par les particules proches comme dans certains modèles (de Boltzmann par exemple). Dans ce sens ce problème de limite est dit *de champ moyen*.

Comme $\hat{\mu}^N$ est solution (faible) de l'équation de Vlasov pour la donnée initiale $\hat{\mu}_0^N$ (cf. [25] ou [104] par exemple), et comme évidemment il en est de même pour f et f_0, le problème de la limite macroscopique se ramène donc à un problème de stabilité des solutions de l'équation de Vlasov (nous verrons dans la partie III de cette introduction que ce n'est plus le cas dans un cadre où l'évolution est stochastique). Une première réponse à cette question de stabilité a été donnée, relativement à la distance d_{BL} qui métrise la topologie faible (étroite) sur $\mathcal{P}(\mathbb{R}^{2d})$, sous la forme suivante :

Théorème 1 (cf. [85], [104]). *Supposons que $\nabla_x V$ soit une fonction sur $\mathbb{R}^d$ bornée par une constante B et lipschitzienne de semi-norme de Lipschitz L. Si f et g sont des solutions de l'équation de Vlasov* (5) *de données initiales respectives f_0 et g_0 dans $\mathcal{P}(\mathbb{R}^{2d})$, alors*

$$d_{BL}(f_t, g_t) \leq e^{c\,t}\, d_{BL}(f_0, g_0)$$

pour tout $t \geq 0$, où c est la constante $(2\max(B,1)+1)\max(L,1)$.

Dans cet énoncé la distance d_{BL} est définie par

$$d_{BL}(\mu,\nu) = \sup_{\|\varphi\|_{lip}\leq 1} \left\{ \int_{\mathbb{R}^{2d}} \varphi(z)\, d\mu(z) - \int_{\mathbb{R}^{2d}} \varphi(z)\, d\nu(z) \right\}$$

où $\|\,.\,\|_{lip}$ désigne la norme de Lipschitz définie par

$$\|\varphi\|_{lip} = \max\left[\sup_z |\varphi(z)|,\, \sup_{w\neq z} \frac{|\varphi(w)-\varphi(z)|}{\|w-z\|_{\ell^1}}\right]$$

et $\|\,.\,\|_{\ell^1}$ désigne la norme sur $\mathbb{R}^{2d}$ définie par $\|(z_1,z_2)\|_{\ell^1} = |z_1| + |z_2|$ où $|z_i|$ est la norme euclidienne de z_i dans $\mathbb{R}^d$.

Sur l'espace $\mathcal{P}_1(\mathbb{R}^{2d})$ des mesures de $\mathcal{P}(\mathbb{R}^{2d})$ de premier moment fini considérons maintenant la distance de Wasserstein W_1 d'ordre 1 définie à partir de la norme $\|.\|_{\ell^1}$ sur $\mathbb{R}^{2d}$. Par la formulation de Kantorovich-Rubinstein on note que $W_1 \geq d_{BL}$.

En adaptant la démonstration proposée dans [104] pour la distance d_{BL}, [111, Problem 14] donne une version du théorème 1 dans le cadre de la distance W_1 pour des données initiales dans $\mathcal{P}_1(\mathbb{R}^{2d})$. Dans le chapitre 1 nous précisons ce résultat sous la forme suivante :

Théorème 2 (cf. Theorem 1.2). *Supposons que $\nabla_x V$ soit une fonction lipschitzienne sur $\mathbb{R}^d$ de semi-norme de Lipschitz L. Si f et g sont des solutions de l'équation de Vlasov* (5) *de données initiales respectives f_0 et g_0 dans $\mathcal{P}_1(\mathbb{R}^{2d})$, alors*

$$W_1(f_t, g_t) \leq e^{ct}\, W_1(f_0, g_0)$$

pour tout $t \geq 0$, où c est la constante $L + \max(L, 1)$.

Ce résultat est donné dans [111] avec la constante $c = 2\,\max(L, 1)$.

Dans le chapitre 1 nous montrons ensuite que la distance de Wasserstein W_2 d'ordre 2, définie sur l'espace $\mathcal{P}_2(\mathbb{R}^{2d})$ des mesures de $\mathcal{P}(\mathbb{R}^{2d})$ de second moment fini, à partir de la norme euclidienne sur $\mathbb{R}^{2d}$, permet de nouveau d'améliorer la constante c intervenant dans les théorèmes 1 et 2. Plus précisément, en utilisant la structure de type différentiel de l'espace $(\mathcal{P}_2(\mathbb{R}^{2d}), W_2)$ évoquée précédemment, nous obtenons de manière formelle le

Théorème 3 (cf. Theorem 1.3). *Supposons que V soit une fonction deux fois différentiable sur $\mathbb{R}^d$, de matrice hessienne $D^2_x V$ telle que $-L\,Id \leq D^2_x V(x) \leq L\,Id$ pour une constante L et tout x dans $\mathbb{R}^d$. Si f et g sont des solutions de l'équation de Vlasov* (5) *de données initiales respectives f_0 et g_0 dans $\mathcal{P}_2(\mathbb{R}^{2d})$ ayant mêmes moyennes en espace et en vitesse, alors*

$$W_2(f_t, g_t) \leq e^{ct}\, W_2(f_0, g_0)$$

pour tout $t \geq 0$, où c est la constante $(L+1)/2$.

Nous voyons sur cet exemple comment l'utilisation de la distance W_2, éventuellement par une preuve plus complexe s'appuyant sur une théorie plus élaborée, permet d'affiner un résultat obtenu par l'utilisation de la distance W_1, peut-être de manière plus simple grâce à la formulation duale de Kantorovich-Rubinstein.

Nous retrouverons ce type d'amélioration dans le cadre de l'approximation particulaire des équations d'Euler, vers laquelle nous nous tournons maintenant.

I. 2. Approximation particulaire des équations d'Euler incompressibles dans le plan (cf. chapitre 2)

Nous venons de voir que le problème de la limite de champ moyen pour l'équation de Vlasov (avec un potentiel lipschitzien) pouvait être ramené à un problème de stabilité des solutions de cette équation ; ce problème a été à son tour résolu de manière quantitative par une propriété de contraction (dans un sens large) de la forme

$$d(f_t, g_t) \leq e^{ct}\, d(f_0, g_0)$$

où d désigne la distance d_{BL}, W_1 ou W_2 et c est une constante dépendant du potentiel et de la distance choisie, la précision de la méthode étant déterminée par le facteur e^{ct}.

Nous considérons maintenant un problème analogue d'approximation particulaire pour les équations d'Euler incompressibles dans le plan.

Ces équations, portant sur la vitesse $u = u(t,x) \in \mathbb{R}^2$ d'un fluide incompressible, non visqueux et recouvrant le plan $\mathbb{R}^2$, s'écrivent

$$\left\{ \begin{array}{rcl} \dfrac{\partial u}{\partial t} + (u \cdot \nabla)\, u & = & -\nabla p \\ \nabla \cdot u & = & 0 \end{array} \right. \qquad t > 0, \quad x \in \mathbb{R}^2 \tag{7}$$

où $\nabla \cdot$ est l'opérateur de divergence sur $\mathbb{R}^2$. La fonction $p = p(t,x)$ désigne le champ scalaire de pression et n'est pas une donnée du problème, mais une inconnue supplémentaire que l'on peut déterminer à partir de u.

Au champ de vitesse $u_t : x \mapsto u(t,x)$ au temps t peut être associé le champ scalaire de vorticité (ou tourbillon) ω_t défini sur $\mathbb{R}^2$ par

$$\omega_t \;=\; \mathrm{rot}\, u_t = \frac{\partial u_t^2}{\partial x_1} - \frac{\partial u_t^1}{\partial x_2}$$

si $u_t = (u_t^1, u_t^2)$ et $x = (x_1, x_2)$ dans $\mathbb{R}^2$. Ce champ mesure le degré de rotation du fluide en chaque point, et d'après (7) est transporté par le champ de vitesse u_t dans le sens où $\omega : t \mapsto \omega_t$ est solution de l'équation

$$\frac{\partial \omega}{\partial t} + u \cdot \nabla \omega \;=\; 0, \qquad t > 0, \quad x \in \mathbb{R}^2. \tag{8}$$

u_t pouvant s'écrire $u_t = K * \omega_t$ où $K = \Big(\dfrac{\partial G}{\partial x_2}, \, - \dfrac{\partial G}{\partial x_1} \Big)$ et $G(x) = -\dfrac{1}{2\,\pi} \ln |x|$ est la solution fondamentale de l'équation de Poisson sur $\mathbb{R}^2$, cette équation (8) s'écrit aussi sous la forme

$$\frac{\partial \omega}{\partial t} \;+\; (K * \omega) \cdot \nabla \omega \;=\; 0, \qquad t > 0, \quad x \in \mathbb{R}^2. \tag{9}$$

C'est cette équation (9), formellement équivalente à (7), que nous considérons désormais. Il s'agit là encore d'une équation de champ moyen dans la mesure où l'évolution de $\omega_t(x)$ en un point x fixé dépend, via le terme de convolution $K * \omega_t(x)$, de la valeur de ω_t en tous les autres points de $\mathbb{R}^2$.

Les solutions considérées sont à chaque instant $t \geq 0$ des mesures de probabilité sur $\mathbb{R}^2$, absolument continues par rapport à la mesure de Lebesgue et à densité bornée presque partout. Dans ce cadre, il existe d'après [76] une unique solution de donnée initiale fixée, dans un sens précisé dans le chapitre 2.

Nous étudions maintenant une approximation particulaire de l'équation (9), appelée *méthode des tourbillons* et définie comme suit.

Soit ω_0^N un profil initial de vorticité de la forme

$$\omega_0^N = \sum_{i=1}^{N} a_i \, \delta_{X_0^{i,N}}$$

où les $X_0^{i,N}$ sont N points de $\mathbb{R}^2$ et les a_i sont N nombres positifs de somme 1. Le terme $a_i\,\delta_{X_0^{i,N}}$ est appelé un tourbillon d'intensité a_i et localisé en $X_0^{i,N}$. Supposant alors que les N points X_0^i évoluent en $X_t^{i,N}$ selon les équations

$$\frac{dX_t^{i,N}}{dt} = \sum_{\substack{j=1\\ j\neq i}}^{N} a_j\, K(X_t^{i,N} - X_t^{j,N}), \qquad 1 \leq i \leq N,$$

la mesure

$$\omega_t^N = \sum_{i=1}^{N} a_i\, \delta_{X_t^{i,N}}$$

peut être interprétée comme une solution faible de l'équation (9) (cf. [76], [97] ou [98]). Notons que cette solution n'entre pas dans le cadre de solutions à densité considéré ci-dessus.

Le problème de l'approximation particulaire pourrait alors s'énoncer ainsi : étant donné un profil initial de vorticité ω_0 sous forme d'une mesure de probabilité sur $\mathbb{R}^2$ à densité bornée presque partout, que l'on suppose bien approché (dans un sens à préciser) par un profil discret ω_0^N, le profil ω_t obtenu par l'équation d'Euler (9) est-il alors toujours bien approché par le profil discret ω_t^N construit ci-dessus ?

Cependant, pour éviter des difficultés dues à la singularité du noyau K en 0, on remplace K par une fonction K_ε de classe $\mathcal{C}^\infty$ de $\mathbb{R}^2$ dans $\mathbb{R}^2$, de divergence nulle, vérifiant $K_\varepsilon(0) = 0$ et $K_\varepsilon(z) = K(z)$ pour $|z| > \varepsilon$ pour un nombre ε positif fixé. On fait alors évoluer les N centres $X_0^{i,N}$ des tourbillons en $X_t^{i,N,\varepsilon}$ par les équations

$$\frac{dX_t^{i,N,\varepsilon}}{dt} = \sum_{j=1}^{N} a_j\, K_\varepsilon(X_t^{i,N,\varepsilon} - X_t^{j,N,\varepsilon}), \qquad 1 \leq i \leq N$$

qui ont maintenant une unique solution, quelles que soient les données initiales $X_0^{i,N}$. Le tourbillon initial $\omega_0^N = \sum_{i=1}^{N} a_i\, \delta_{X_0^{i,N}}$ évolue alors en $\omega_t^{N,\varepsilon} = \sum_{i=1}^{N} a_i\, \delta_{X_t^{i,N,\varepsilon}}$, solution faible de l'équation

$$\frac{\partial \mu}{\partial t} + (K_\varepsilon * \mu) \cdot \nabla \mu = 0, \qquad t > 0, \quad x \in \mathbb{R}^2. \tag{10}$$

Etant donné $T \geq 0$, nous voulons alors savoir pour quels noyaux K_ε et donnée initiale ω_0^N la mesure $\omega_t^{N,\varepsilon}$ reste proche de la solution ω_t préalablement fixée, et ceci sur tout l'intervalle $[0, T]$.

Ce problème est énoncé de manière précise et étudié dans [76]. En particulier l'écart entre les mesures de probabilité considérées y est mesuré au moyen de la distance de Wasserstein $W_{1,d}$ d'ordre 1 construite sur $\mathcal{P}(\mathbb{R}^2)$ à partir de la distance $d(x,y) = \min(|x-y|, 1)$ équivalente sur $\mathbb{R}^2$ à la distance euclidienne $|x-y|$. Comme d_{BL}, cette distance métrise la topologie faible (étroite) sur $\mathcal{P}(\mathbb{R}^2)$. Le résultat obtenu s'énonce ainsi :

Théorème 4 (cf. [76]). *Supposons que le noyau K_ε soit une fonction bornée par la constante B_ε et lipschitzienne de semi-norme de Lipschitz L_ε. Alors, avec les notations introduites ci-dessus, pour tout $T \geq 0$ on a*

$$\sup_{0 \leq t \leq T} W_{1,d}(\omega_t, \omega_t^{N,\varepsilon(N)}) \longrightarrow 0 \quad \textit{quand} \quad N \longrightarrow +\infty$$

pour toute suite $\varepsilon(N)$ telle que

$$\exp\big(c_{\varepsilon(N)}(T + e^{c_{\varepsilon(N)}\, T})\big) W_{1,d}(\omega_0, \omega_0^N) \longrightarrow 0 \quad \textit{quand} \quad N \longrightarrow +\infty$$

où $c_\varepsilon = \max(2\, B_\varepsilon, L_\varepsilon)$.

L'idée de la preuve de ce résultat donnée dans [76] est d'introduire une mesure ω_t^ε proche à la fois de ω_t et de $\omega_t^{N,\varepsilon}$. Pour cela on prend pour ω_t^ε la solution de (10) pour la donnée initiale ω_0. Comme $\omega_t^{N,\varepsilon}$ en est aussi une solution, un résultat général de stabilité assure que

$$W_{1,d}(\omega_t^\varepsilon, \omega_t^{N,\varepsilon}) \leq \exp\big(c_\varepsilon(t + e^{c_\varepsilon\, t})\big) W_{1,d}(\omega_0, \omega_0^N) \tag{11}$$

pour tous $\varepsilon > 0$ et $t \geq 0$.

D'autre part le fait que ω_t et ω_t^ε aient même donnée initiale ω_0 entraîne que

$$\sup_{0 \leq t \leq T} W_{1,d}(\omega_t, \omega_t^\varepsilon) \longrightarrow 0 \quad \text{quand} \quad \varepsilon \longrightarrow 0.$$

Le résultat découle alors de ces deux points par inégalité triangulaire.

Ceci signifie que, étant donnée la suite de données initiales $(\omega_0^N)_N$ convergeant vers ω_0, le paramètre $\varepsilon = \varepsilon(N)$ de troncature du noyau ne doit pas tendre trop vite vers 0, sans quoi $\exp\big(c_\varepsilon(T + e^{c_\varepsilon\, T})\big)$ tendrait trop vite vers l'infini.

Notons de nouveau que la précision de la méthode dépend de la manière dont la constante intervenant dans (11) croît avec t mais aussi avec ε. Il est donc intéressant de voir si cette constante peut être améliorée de manière simple. Notant W_1 et W_2 les distances de Wasserstein d'ordre 1 et 2 définies sur $\mathcal{P}_1(\mathbb{R}^2)$ et $\mathcal{P}_2(\mathbb{R}^2)$ à partir de la distance euclidienne $|x - y|$ sur $\mathbb{R}^2$, nous pouvons améliorer cette constante d'un facteur exponentiel :

Théorème 5 (cf. Theorem 2.2). *Soient ω_t^1 et ω_t^2 deux solutions de (10) de données initiales ω_0^1 et ω_0^2 dans $\mathcal{P}_1(\mathbb{R}^2)$. Alors, avec les notations précédentes,*

$$W_1(\omega_t^1, \omega_t^2) \leq e^{2\, L_\varepsilon\, t}\, W_1(\omega_0^1, \omega_0^2)$$

pour tout $t \geq 0$.

Si de plus le noyau K_ε est une fonction impaire et si les données initiales ω_0^1 et ω_0^2 sont dans $\mathcal{P}_2(\mathbb{R}^2)$ et ont même centre de masse, alors

$$W_2(\omega_t^1, \omega_t^2) \leq e^{L_\varepsilon\, t}\, W_2(\omega_0^1, \omega_0^2)$$

pour tout $t \geq 0$.

Ces résultats sont démontrés dans le chapitre 2 en adaptant des techniques utilisées dans le chapitre 1 pour l'équation de Vlasov.

Obtenir de telles estimations pour l'équation d'Euler originale (9) semble une tâche bien plus difficile. Cependant, dans le cas où l'équation est posée sur un domaine borné D de $\mathbb{R}^2$ assez régulier, et pour des données initiales ω_0^1 et ω_0^2 à densité bornée et telles que $W_1(\omega_0^1, \omega_0^2) \leq 1$, nous montrons (cf. **Theorem 2.5**) l'estimation

$$W_1(\omega_t^1, \omega_t^2) \leq e^{1-\exp(-c\,t)}\, W_1(\omega_0^1, \omega_0^2)^{\exp(-c\,t)}$$

pour tout $t \leq \ln\big(1 - \ln W_1(\omega_0^1, \omega_0^2)\big)/c$, où c est une constante dépendant de D, $\|\omega_0^1\|_{L^\infty(D)}$ et $\|\omega_0^2\|_{L^\infty(D)}$.

I.3. Métriques contractantes pour des lois de conservation scalaires (cf. chapitre 3*)

Dans le chapitre 3 nous établissons une propriété de contraction en distance de Wasserstein pour des solutions croissantes de lois de conservation scalaires, ou plus précisément pour leurs dérivées spatiales.

Etant donnée une fonction de flux f de $\mathbb{R}$ dans $\mathbb{R}$ localement lipschitzienne, considérons plus précisément la loi de conservation scalaire

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = 0\,, \qquad t > 0,\ x \in \mathbb{R} \tag{12}$$

d'inconnue $u = u(t,x)$ à valeurs réelles et de donnée initiale u_0 dans $L^\infty(\mathbb{R})$.

Il a été établi qu'une telle équation peut ne pas admettre de solution classique, mais par contre admettre une infinité de solutions distributions (cf. [41] et [101] par exemple pour l'équation de Burgers).

ll s'agit donc de choisir une solution parmi toutes ces solutions distributions : c'est ce à quoi pourvoit la notion de *solution entropique* (ou admissible), définie par exemple dans [101]. En particulier toute solution classique est une solution entropique. Cette notion de solution entropique est bien adaptée à l'équation considérée dans la mesure où, comme l'a montré S. Kružkov [65], il existe une unique solution entropique de (12) dans $C([0,+\infty[, L^1_{loc}(\mathbb{R}))$, localement bornée et de donnée initiale u_0 dans $L^\infty(\mathbb{R})$.

Comme précédemment, pour tout instant $t > 0$ nous notons u_t la fonction définie dans $\mathbb{R}$ par $u_t(x) = u(t,x)$. Parmi les nombreuses propriétés de ces solutions, mentionnons la propriété suivante de contraction en norme L^1 : si u_0 et $\tilde{u}_0$ sont deux fonctions de $L^\infty(\mathbb{R})$ telles que $u_0 - \tilde{u}_0$ appartienne à $L^1(\mathbb{R})$, alors $u_t - \tilde{u}_t$ est dans $L^1(\mathbb{R})$ pour tout t, avec $\|u_t - \tilde{u}_t\|_{L^1(\mathbb{R})} \leq \|u_0 - \tilde{u}_0\|_{L^1(\mathbb{R})}$.

Pour une certaine classe de solutions nous nous proposons d'étendre cette propriété de contraction à une famille de distances mesurant l'écart entre deux solutions.

* Le chapitre 3 reprend et précise le travail [21] écrit en collaboration avec Y. Brenier et G. Loeper.

Considérons pour cela l'ensemble $\mathcal{U}$ des fonctions croissantes de $\mathbb{R}$ dans $\mathbb{R}$, continues à droite et admettant les limites 0 et 1 en $-\infty$ et $+\infty$. Des lois de conservation scalaires avec des données initiales dans $\mathcal{U}$ ont été utilisées dans [30] dans l'étude de gaz sans pression formé de particules collantes (sticky particles).

Cet ensemble $\mathcal{U}$ est en fait l'ensemble des fonctions de répartition de variables aléatoires réelles, en bijection avec l'ensemble des mesures de probabilité sur $\mathbb{R}$ par l'application qui à une fonction v de $\mathcal{U}$ associe sa dérivée v' au sens des distributions. De plus la relation (3) assure que pour deux fonctions v et $\tilde{v}$ de $\mathcal{U}$

$$\|v - \tilde{v}\|_{L^1(\mathbb{R})} = W_1(v', \tilde{v}').$$

L'ensemble $\mathcal{U}$ étant préservé par la loi de conservation (12) dans le sens où toute solution entropique u de donnée initiale u_0 dans $\mathcal{U}$ est telle que u_t appartient à $\mathcal{U}$ pour tout t, on a en particulier la relation

$$\|u_t - \tilde{u}_t\|_{L^1(\mathbb{R})} = W_1\Big(\frac{\partial u_t}{\partial x}, \frac{\partial \tilde{u}_t}{\partial x}\Big)$$

si u et $\tilde{u}$ sont deux solutions de données initiales dans $\mathcal{U}$, et la propriété de contraction en norme L^1 s'écrit alors

$$W_1\Big(\frac{\partial u_t}{\partial x}, \frac{\partial \tilde{u}_t}{\partial x}\Big) \leq W_1\Big(\frac{\partial u_0}{\partial x}, \frac{\partial \tilde{u}_0}{\partial x}\Big).$$

Comme nous le montrerons dans le chapitre 3, cette propriété se généralise à l'ordre p sous la forme suivante :

Théorème 6 (cf. Theorem 3.6). *Si u et $\tilde{u}$ sont deux solutions entropiques de l'équation* (12) *de données initiales respectives u_0 et $\tilde{u}_0$ dans $\mathcal{U}$, alors*

$$W_p\Big(\frac{\partial u_t}{\partial x}, \frac{\partial \tilde{u}_t}{\partial x}\Big) \leq W_p\Big(\frac{\partial u_0}{\partial x}, \frac{\partial \tilde{u}_0}{\partial x}\Big)$$

pour tous $p \geq 1$ et $t \geq 0$.

De ce résultat nous déduisons que la distance W_p entre les dérivées spatiales de deux solutions entropiques est une fonction décroissante du temps. Nous pouvons alors envisager de prolonger ce résultat dans différentes directions : par exemple montrer que cette distance est conservée par l'équation pour certaines données initiales et fonctions de flux ; étendre cette propriété de contraction à des lois de conservation visqueuses ou à des fonctions de flux différentes ; considérer une situation analogue en dimension d'espace supérieure ; enfin déterminer la limite de la distance quand le temps t tend vers l'infini, et en particulier donner des conditions sur les données initiales pour que cette distance décroisse vers 0.

Nous apportons une réponse à la première question de conservation sous la forme suivante :

Théorème 7 (cf. Theorem 3.7). *Supposons que le flux f soit de classe $\mathcal{C}^1$. Si deux solutions classiques u et $\tilde{u}$ de l'équation* (12) *de données initiales respectives u_0 et $\tilde{u}_0$ dans $\mathcal{U}$ sont strictement croissantes en x pour tout $t \geq 0$, alors*

$$W_p\Big(\frac{\partial u_t}{\partial x}, \frac{\partial \tilde{u}_t}{\partial x}\Big) = W_p\Big(\frac{\partial u_0}{\partial x}, \frac{\partial \tilde{u}_0}{\partial x}\Big)$$

pour tous $t \geq 0$ *et* $p \geq 1$. *En particulier*

$$\|u_t - \tilde{u}_t\|_{L^1(\mathbb{R})} = \|u_0 - \tilde{u}_0\|_{L^1(\mathbb{R})}$$

pour tout $t \geq 0$.

Les hypothèses de ce théorème sont vérifiées en particulier dès que f est convexe et de classe $\mathcal{C}^2$, et u_0 et $\tilde{u}_0$ sont strictement croissantes et de classe $\mathcal{C}^1$ (cf. **Corollary 3.11**).

Notons que cette question de conservation de la distance en norme L^1 entre deux solutions entropiques est abordée par C. Dafermos [41, Section 11.8] dans le cadre (différent) de solutions dans $L^1(\mathbb{R})$ et de flux strictement convexes.

La preuve du théorème 7 est fondée sur les deux remarques suivantes : d'une part, si X_0 est l'inverse de u_0, la méthode des caractéristiques assure qu'en tout temps $t \geq 0$ l'inverse X_t de u_t est donné par

$$X_t(w) = X_0(w) + t\, f'(w)$$

et d'autre part, avec des notations analogues pour $\tilde{u}$, la relation (2) assure l'égalité

$$W_p^p\Big(\frac{\partial u_t}{\partial x}, \frac{\partial \tilde{u}_t}{\partial x}\Big) = \int_0^1 |X_t(w) - \tilde{X}_t(w)|^p\, dw \tag{13}$$

pour tout $t \geq 0$, et en particulier pour $t = 0$.

Cette idée est à la base de la démonstration du théorème 6, mais dans le cadre général de fonctions de $\mathcal{U}$ nous ne pouvons plus considérer leurs inverses, mais plutôt leurs pseudo-inverses. Faisons alors évoluer le pseudo-inverse X_0 de u_0 suivant la méthode des caractéristiques, sur un intervalle de temps h, en

$$X_h(w) = X_0(w) + h\, f'(w).$$

X_h n'étant plus a priori croissante, nous ne pouvons considérer son inverse ni même son pseudo-inverse. Nous introduisons alors le pseudo-inverse $T_h u_0$ de son réarrangement monotone : c'est la fonction de répartition de X_h, et est donc un élément de $\mathcal{U}$. Nous retrouvons ici la méthode de transport-écroulement développée par Y. Brenier dans [26] et [27] et qui permet de construire une solution approchée de (12).

Contrairement à l'égalité (13) du cas classique du théorème 7, nous avons seulement, dans le cas général du théorème 6, la majoration

$$W_p^p\Big(\frac{\partial}{\partial x}(T_h u_0), \frac{\partial}{\partial x}(T_h \tilde{u}_0)\Big) \leq \int_0^1 |X_h(w) - \tilde{X}_h(w)|^p\, dw \ \left(= W_p^p\Big(\frac{\partial u_0}{\partial x}, \frac{\partial \tilde{u}_0}{\partial x}\Big)\right). \tag{14}$$

Par itération de l'opérateur T_h nous définissons à partir de u_0 et $\tilde{u}_0$ deux solutions approchées de (12) vérifiant en tout temps la propriété de contraction cherchée (d'après (14)). Montrant que ces solutions approchées convergent respectivement vers u et $\tilde{u}$ dans un sens convenable quand h tend vers 0, nous en déduisons alors le théorème 6.

Ce même théorème 6 peut être étendu aux solutions (distributions) de lois de conservation visqueuses de la forme

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = \nu \, \frac{\partial^2 u}{\partial x^2}, \qquad t > 0, \; x \in \mathbb{R}, \tag{15}$$

où ν est un nombre strictement positif fixé. Pour cela nous complétons le schéma développé pour la loi de conservation inviscide de la manière suivante : à une donnée initiale u_0 nous associons la fonction $T_h u_0$ précédemment définie, et qui rend compte du terme $\partial f(u)/\partial x$ de l'équation ; puis nous faisons évoluer $T_h u_0$ en $\mathcal{T}_h u_0$ selon l'équation de la chaleur avec viscosité ν sur l'intervalle de temps h, rendant ainsi compte du terme de diffusion $\nu \, \partial^2 u / \partial x^2$. Par itération de cet opérateur $\mathcal{T}_h$ nous pouvons alors obtenir l'analogue du théorème 6 dans le cas visqueux (cf. **Theorem 3.28**).

Plus généralement nous pouvons nous intéresser aux solutions de lois de conservation de fonctions de flux différentes. Supposons pas exemple que u et $\tilde{u}$ soient des solutions de

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = \nu \, \frac{\partial^2 u}{\partial x^2}, \qquad \frac{\partial \tilde{u}}{\partial t} + \frac{\partial}{\partial x} \tilde{f}(\tilde{u}) = \nu \, \frac{\partial^2 \tilde{u}}{\partial x^2} \qquad t > 0, \quad x \in \mathbb{R} \tag{16}$$

de données initiales respectives u_0 et $\tilde{u}_0$ dans $\mathcal{U}$, où $\nu \geq 0$. Dans le cas inviscide où $\nu = 0$, Y. Brenier [29] obtient la relation

$$W_2\Big(\frac{\partial u_t}{\partial x}, \frac{\partial \tilde{u}_t}{\partial x}\Big) \leq W_2\Big(\frac{\partial u_0}{\partial x}, \frac{\partial \tilde{u}_0}{\partial x}\Big) + t \, \|f' - \tilde{f}'\|_{L^2(]0,1[)}$$

pour $t \geq 0$. Nous étendons cette propriété au cas visqueux et à tout $p \geq 1$ sous la forme suivante :

Théorème 8 (cf. Theorem 3.36). *Soient $\nu \geq 0$, f et $\tilde{f}$ deux fonctions localement lipschitziennes. Si u et $\tilde{u}$ sont des solutions (entropiques si $\nu = 0$ et distributions si $\nu > 0$) des équations* (16) *de données initiales u_0 et $\tilde{u}_0$ dans $\mathcal{U}$, alors*

$$W_p\Big(\frac{\partial u_t}{\partial x}, \frac{\partial \tilde{u}_t}{\partial x}\Big) \leq W_p\Big(\frac{\partial u_0}{\partial x}, \frac{\partial \tilde{u}_0}{\partial x}\Big) + t \, \|f' - \tilde{f}'\|_{L^p(]0,1[)}$$

pour tous $t \geq 0$ et $p \geq 1$.

Ce résultat, qui contient bien entendu le théorème 6 et son équivalent visqueux, sera en fait démontré à l'aide d'estimations servant déjà à montrer ces résultats.

Nous avons donc établi que la distance entre les dérivées spatiales de deux solutions des lois de conservation (12) et (15), à données initiales dans $\mathcal{U}$, est une fonction décroissante du temps. Une question serait alors de déterminer la limite de cette fonction, et si possible de montrer que cette limite est égale à 0 pour certaines données initiales. Ceci est envisageable dans la mesure où des résultats de convergence de solutions vers des profils particuliers ont été obtenus dans le cadre L^1 (cf. [102]) ; citons en particulier l'étude du comportement asymptotique de solutions de l'équation de Burgers avec viscosité menée par M. Di Francesco et P. A. Markowich [45] et déjà citée dans cette introduction.

On pourrait également étudier les mêmes problèmes pour d'autres types d'équations, comme par exemple l'équation de Rosenau

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = K * u - u\,, \qquad t > 0,\ x \in \mathbb{R},$$

où $K(x) = e^{-|x|}/2$, qui est un modèle simplifié de gaz radiatif (cf. [102]).

Parmi d'autres généralisations envisageables on pourrait considérer des lois de conservation scalaires sur $[0, +\infty[\times \mathbb{R}^d$ avec $d \geq 1$ et s'intéresser à des solutions croissantes en chacune des variables d'espace, et tendant vers 0 (resp. 1) quand chaque variable d'espace tend vers $-\infty$ (resp. $+\infty$). De telles fonctions peuvent encore s'interpréter comme des fonctions de répartition de variables aléatoires à valeurs dans $\mathbb{R}^d$, et dans la mesure où $\|u_t - \tilde{u}_t\|_{L^1(\mathbb{R}^d)}$ est toujours une fonction décroissante du temps si u et $\tilde{u}$ sont deux telles solutions, on peut dans ce cadre multidimensionnel se poser les mêmes questions que dans le cadre monodimensionnel précédent ; par contre on ne peut plus exprimer directement les distances de Wasserstein en termes de pseudo-inverses.

II - Inégalités de concentration et de transport

Dans l'étude des sytèmes de particules en interaction que nous présenterons dans la partie III de cette introduction, nous établirons en particulier des inégalités de déviation de la forme

$$\mathbb{P}\left[\sup_{[\varphi]_{lip}\leq 1} \left\{\frac{1}{N}\sum_{i=1}^{N} \varphi(X_t^{i,N}) - \int_{\mathbb{R}^d} \varphi(x)\, d\mu_t(x)\right\} > \varepsilon\right] \leq e^{-KN\varepsilon^2}$$

pour tous $\varepsilon > 0$ et N assez grand, où, pour tout instant $t \geq 0$, μ_t est une mesure de probabilité sur $\mathbb{R}^d$ décrivant l'état du système au niveau macroscopique, et les $X_t^{i,N}$ donnent l'état dans $\mathbb{R}^d$ de N particules en interaction ; comme précédemment $[\varphi]_{lip}$ désigne la semi-norme de Lipschitz d'une fonction φ de $\mathbb{R}^d$ dans $\mathbb{R}$, définie à partir de la distance euclidienne sur $\mathbb{R}^d$. Pour obtenir une telle inégalité nous introduirons N variables aléatoires $Y_t^{i,N}$ pour $1 \leq i \leq N$, proches des $X_t^{i,N}$ dans un certain sens, indépendantes et de loi commune μ_t. Oubliant ce problème particulier, il s'agira alors de montrer une inégalité générale de la forme

$$\mathbb{P}\left[\sup_{[\varphi]_{lip}\leq 1} \left\{\frac{1}{N}\sum_{i=1}^{N} \varphi(Y^i) - \int_{\mathbb{R}^d} \varphi(x)\, d\mu(x)\right\} > \varepsilon\right] \leq e^{-KN\varepsilon^2}, \tag{17}$$

et dans un premier temps l'inégalité

$$\sup_{[\varphi]_{lip}\leq 1} \mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} \varphi(Y^i) - \int_{\mathbb{R}^d} \varphi(x)\, d\mu(x) > \varepsilon\right] \leq e^{-KN\varepsilon^2}$$

où μ est un mesure de probabilité sur $\mathbb{R}^d$ et Y^i pour $1 \leq i \leq N$ sont N variables aléatoires indépendantes de loi μ.

C'est ce type d'inégalités de concentration, ou de déviation, que nous considérons dans cette partie, en en donnant en particulier des conditions nécessaires ou suffisantes pour qu'elles soient vérifiées, en étudiant leurs liens avec les inégalités de transport et de Sobolev logarithmiques et en cherchant dans quelle mesure elles peuvent s'étendre à certaines familles de variables Y^i dépendantes.

II.1. Inégalités de concentration gaussienne (cf. chapitre 4*)

Considérons tout d'abord le cas d'une mesure μ sur $\mathbb{R}$. Sous quelles conditions sur μ sommes-nous certains qu'une inégalité telle que

$$\sup_{[\varphi]_{lip}\leq 1} \mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N}\varphi(Y^i) - \int_{\mathbb{R}}\varphi(x)\,d\mu(x) > \varepsilon\right] \leq e^{-\frac{\lambda}{2}N\varepsilon^2} \tag{18}$$

et en premier lieu

$$\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N}Y^i - \int_{\mathbb{R}}x\,d\mu(x) > \varepsilon\right] \leq e^{-\frac{\lambda}{2}N\varepsilon^2}$$

soit vérifiée pour une certaine constante $\lambda > 0$, tout $\varepsilon > 0$, tout $N \geq 1$ et tout échantillon $(Y^i)_{1\leq i\leq N}$ de N variables aléatoires indépendantes et de loi μ ?

Si la mesure μ est de variance finie, le théorème central limite assure que l'inégalité (18) est vérifiée pour une constante λ dépendant de la variance de μ, mais seulement asymptotiquement quand N tend vers l'infini. Pour qu'elle le soit pour tout N, nous devons ajouter des conditions sur μ. Par exemple, si μ a son support dans un segment $[a, b]$ de $\mathbb{R}$, alors l'inégalité de Hoeffding (cf. [68, Section 1.6] ou [81, Section 1.2] par exemple) assure que

$$\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N}Y^i - \int_{\mathbb{R}}x\,d\mu(x) > \varepsilon\right] \leq e^{-\frac{2N\varepsilon^2}{(b-a)^2}} \tag{19}$$

puis que

$$\sup_{[\varphi]_{lip}\leq 1} \mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N}\varphi(Y^i) - \int_{\mathbb{R}}\varphi(x)\,d\mu(x) > \varepsilon\right] \leq e^{-\frac{2N\varepsilon^2}{(b-a)^2}} \tag{20}$$

pour tous $\varepsilon > 0$ et $N \geq 1$, ce qui répond à la question.

Plus généralement, (18) est vérifiée pour tout N si l'inégalité

$$\sup_{[\varphi]_{lip}\leq 1}\int_{\mathbb{R}} e^{t\,(\varphi(x)-\int_{\mathbb{R}}\varphi(y)\,d\mu(y))}\,d\mu(x) \leq e^{\frac{t^2}{2\lambda}},$$

est vérifiée pour tout réel t. Cette condition est en effet suffisante comme l'assure l'inégalité de Chebyschev exponentielle ; elle est aussi nécessaire (cf. [56, Proposition VI. 48] par exemple

* Le chapitre 4 correspond en grande partie à l'article [23] écrit en collaboration avec C. Villani.

dans un cadre plus général). Notons que cette condition est vérifiée avec $\lambda = 4/(b-a)^2$ pour toute mesure μ dont le support est inclus dans le segment $[a, b]$, ce qui permet de retrouver (19) et (20).

De manière générale, si (X, d) est un espace métrique complet et séparable, nous dirons qu'une mesure borélienne de probabilité μ sur X vérifie une *inégalité de concentration gaussienne* de constante $\lambda > 0$, notée $CG(\lambda)$, si l'inégalité

$$\sup_{[\varphi]_{lip} \leq 1} \int_X e^{t\,(\varphi(x) - \int_X \varphi(y)\, d\mu(y))}\, d\mu(x) \leq e^{\frac{t^2}{2\lambda}} \tag{21}$$

est vérifiée pour tout réel t. Par exemple (21) est vérifiée par la mesure gaussienne standard sur $\mathbb{R}$ avec $\lambda = 1$, avec égalité pour $\varphi(x) = x$. En termes d'inégalité de déviation pour les fonctions lipschitziennes, (21) est une condition suffisante pour que l'inégalité de déviation normale (gaussienne)

$$\sup_{[\varphi]_{lip} \leq 1} \mu\Big[\varphi - \int_X \varphi(x)\, d\mu(x) > \varepsilon\Big] \leq e^{-\frac{\lambda}{2}\varepsilon^2},$$

qui est (18) avec $N = 1$, soit vérifiée pour tout $\varepsilon > 0$ (cf. M. Ledoux [68, Chapter 1] pour une étude de cette propriété).

Cette inégalité $CG(\lambda)$ admet une formulation duale équivalente $T_1(\lambda)$ que nous définissons maintenant. Pour cela, soit W_1 la distance de Wasserstein d'ordre 1 définie à partir de la distance d sur X, et, étant données deux mesures boréliennes de probabilité μ et ν sur X, soit $H(\nu|\mu)$ l'*entropie relative* (ou information de Kullback) de ν par rapport à μ définie par

$$H(\nu|\mu) = \int_X \frac{d\nu}{d\mu}(x)\, \ln \frac{d\nu}{d\mu}(x)\, d\mu(x)$$

si ν est absolument continue par rapport à μ, de dérivée de Radon-Nikodym $\dfrac{d\nu}{d\mu}$, et par $H(\nu|\mu) = +\infty$ sinon.

Avec ces notations, nous dirons qu'une mesure borélienne de probabilité μ sur (X, d) vérifie une *inégalité de transport d'ordre* 1 de constante $\lambda > 0$, notée $T_1(\lambda)$, si l'inégalité

$$W_1(\nu, \mu) \leq \sqrt{\frac{2}{\lambda}\, H(\nu|\mu)} \tag{22}$$

est satisfaite pour toute mesure ν sur X et nous dirons que μ vérifie une inégalité T_1 si elle satisfait $T_1(\lambda)$ pour un $\lambda > 0$.

En utilisant les formulations duales de la distance W_1 (de Kantorovich-Rubinstein) et de l'entropie (cf. [43, Lemma 6. 2. 13] par exemple), S. Bobkov et F. Götze [17] ont alors montré qu'une mesure μ vérifie l'inégalité $CG(\lambda)$ si et seulement si elle vérifie $T_1(\lambda)$.

Si l'espace X est muni de la distance $d(x, y) = \mathbf{1}_{x \neq y}$, toute mesure μ sur X vérifie l'inégalité $T_1(4)$ puisque (22) se réduit alors à l'inégalité de Csiszár-Kullback-Pinsker

$$\|\nu - \mu\|_{TV} \leq \sqrt{2\, H(\nu|\mu)}$$

qui elle est vraie pour toutes mesures ν et μ sur X (cf. par exemple [5, Théorème 8.2.7], [43, Exercise 6.2.17] et la preuve du théorème 4.1 dans le chapitre 4, avec $\varphi \equiv 1$, pour trois démonstrations différentes). Notons que cette inégalité est liée à celle de Hoeffding (19), comme l'attestent par exemple les démonstrations de ces deux inégalités données dans [81] et fondées sur le même lemme, et nous avons vu que (19) est un cas simple de l'inégalité de déviation (18) considérée dans ce paragraphe.

Nous venons de noter que l'inégalité de Csiszár-Kullback-Pinsker est vraie pour toutes mesures μ et ν. Il n'en est pas de même des inégalités équivalentes $CG(\lambda)$ et $T_1(\lambda)$, comme l'atteste la caractérisation suivante :

Théorème 9 (cf. [46] et Corollary 4.6). *Une mesure de probabilité borélienne μ sur X vérifie une inégalité T_1 (ou de concentration gaussienne) si et seulement si elle admet un moment exponentiel carré, au sens où il existe $a > 0$ et $x_0 \in X$ tels que* $\displaystyle\int_X e^{ad(x,x_0)^2}\, d\mu(x)$ *soit fini.*

Plus précisément, si μ vérifie $T_1(\lambda)$ (ou $CG(\lambda)$), alors $\displaystyle\int_X e^{ad(x,x_0)^2}\, d\mu(x)$ *est fini pour tout $a < \lambda/2$ et $x_0 \in X$.*

Inversement, s'il existe $a > 0$ et $x_0 \in X$ tels que $\displaystyle\int_X e^{ad(x,x_0)^2}\, d\mu(x)$ *soit fini, alors μ vérifie $T_1(\lambda)$ (ou $CG(\lambda)$) avec* $\displaystyle\lambda = \sup_{x_0 \in X,\, a>0} \left(\frac{1}{a}\Big(1 + \ln \int_X e^{ad(x_0,x)^2}\, d\mu(x)\Big)\right)^{-1}$.

La condition nécessaire est donnée dans [46] (cf. aussi [68, Proposition 1.9] et [5, Section 7.2]), alors que la condition suffisante est démontrée dans [46, Theorem 2.3] dans la formulation $CG(\lambda)$ et précisée dans le chapitre 4 dans la formulation $T_1(\lambda)$ (cf. **Corollary 4.6**) ou $CG(\lambda)$ (cf. **Theorem 4.18**), où on obtient l'expression de la constante λ.

Cette caractérisation peut s'avérer pratique pour des applications dans lesquelles la dimension de l'espace ne joue pas un rôle prépondérant. Par exemple, dans l'étude de systèmes de particules en interaction présentée dans la partie III, montrer qu'à chaque instant t la loi μ_t sur l'espace des phases $\mathbb{R}^d$ vérifie une inégalité T_1 reviendra à montrer l'existence d'un moment exponentiel carré pour la loi μ_t, c'est-à-dire encore à vérifier la propagation d'un tel moment par l'équation aux dérivées partielles satisfaite par μ_t. Cette caractérisation assure également le résultat de perturbation suivant : si h est une fonction bornée, alors la mesure $h\,\mu$ normalisée vérifie une inégalité T_1 s'il en est de même de μ.

L'inégalité de concentration gaussienne $CG(\lambda)$ est liée au phénomène de concentration de la mesure (présenté par M. Ledoux dans [67] et [68] par exemple) sous la forme suivante : si μ vérifie une inégalité $CG(\lambda)$, alors pour tout borélien A de X tel que $\mu[A] \geq 1/2$ et tout $r > 0$ on a

$$\mu[A^r] \geq 1 - e^{-\frac{\lambda}{8}r^2} \tag{23}$$

où $A^r = \{x \in X; d(x, A) \leq r\}$ (cf. [68, Proposition 1.7] par exemple). Autrement dit la mesure de A^r devient rapidement proche de 1 (de manière exponentielle) quand r tend vers l'infini, ce qui semble naturel puisque la queue de la distribution μ est exponentiellement petite d'après le théorème 9.

Avant que S. Bobkov et F. Götze [17] aient montré l'équivalence des inégalités $CG(\lambda)$ et $T_1(\lambda)$, K. Marton [77] (cf. aussi [106]) avait montré que (23) est vérifiée pour tous A et r tels que $\mu(A) \geq 1/2$ et $r \geq 2\sqrt{-2 \ln \mu[A]}$ sous la condition que μ vérifie une inégalité $T_1(\lambda)$.

Ayant ainsi caractérisé les mesures vérifiant l'inégalité (18), nous montrerons dans le paragraphe suivant dans le cadre des inégalités de transport, comment obtenir l'inégalité plus forte (17), cette fois sous une condition sur la taille N de l'échantillon.

II.2. Inégalités de transport (cf. chapitres 4, 6* et 7)

Etendant la définition de l'inégalité T_1 à tout $p \geq 1$, nous dirons qu'une mesure borélienne de probabilité sur X satisfait une inégalité $T_p(\lambda)$, dite *inégalité de transport* ou *de Talagrand* d'ordre p et de constante λ, si

$$W_p(\nu, \mu) \leq \sqrt{\frac{2}{\lambda} H(\nu|\mu)}$$

pour toute mesure ν sur X, et qu'elle vérifie une inégalité T_p si elle satisfait $T_p(\lambda)$ pour un $\lambda > 0$.

Ces inégalités T_p deviennent de plus en plus fortes quand p grandit puisque $W_p \leq W_{p'}$ pour $1 \leq p \leq p'$ d'après l'inégalité de Hölder. Les distances W_1 et W_2 étant les distances de Wasserstein les plus utilisées, comme nous l'avons déjà noté, il en est de même des inégalités T_1 et T_2.

Alors que l'inégalité T_1 peut être caractérisée simplement (comme dans le théorème 9 par exemple), il semble qu'il n'en soit pas de même pour T_2. Elle admet cependant une formulation duale analogue à (21) (cf. [17]). D'autre part, par exemple sur $\mathbb{R}^d$ muni de la distance euclidienne, l'inégalité $T_2(\lambda)$ implique l'inégalité de Poincaré

$$\int_{\mathbb{R}^d} \left| f(x) - \int_{\mathbb{R}^d} f(y)\, d\mu(y) \right|^2 d\mu(x) \leq \frac{1}{\lambda} \int_{\mathbb{R}^d} |\nabla f|^2(x)\, d\mu(x)$$

pour toute fonction f sur $\mathbb{R}^d$ de classe $\mathcal{C}^\infty$ à support compact ; inversement toute mesure μ sur $\mathbb{R}^d$ de la forme $d\mu(x) = e^{-V(x)}\, dx$, où V est une fonction deux fois différentiable sur $\mathbb{R}^d$ telle que $D^2V(x) \geq \lambda\, Id$ pour un $\lambda > 0$ et tout $x \in \mathbb{R}^d$, vérifie $T_2(\lambda)$ (cf. [14], [18], [88]), et plus généralement toute mesure vérifiant une inégalité de Sobolev logarithmique satisfait une inégalité T_2 comme nous le verrons dans le paragraphe II.3.

Nous avons noté que T_2 implique T_1 ; la réciproque est fausse puisqu'il existe des mesures sur $\mathbb{R}$ à support compact, donc vérifiant une inégalité T_1, mais ne satisfaisant pas d'inégalité de Poincaré, et donc a fortiori pas d'inégalité T_2.

Une propriété remarquable de l'inégalité T_2 est la propriété de tensorisation. M. Talagrand [106], après avoir établi que la mesure gaussienne sur $\mathbb{R}$ vérifie $T_2(1)$, en déduit alors qu'il

* Le chapitre 6 reprend l'article [22] écrit en collaboration avec A. Guillin et C. Villani.

en est de même pour la mesure gaussienne sur $\mathbb{R}^N$ avec N quelconque. Plus généralement, si, pour $1 \leq i \leq N$, μ_i est une mesure sur X vérifiant $T_2(\lambda)$, alors il en est de même (en particulier avec la même constante) pour la mesure produit $\mu_1 \otimes \cdots \otimes \mu_N$ sur l'espace X^N muni de la distance $d_{\ell^2}(x,y) = \Big(\sum_{i=1}^N d(x_i,y_i)^2\Big)^{1/2}$ (cf. [68, Proposition 6.3] par exemple).

Dans le paragraphe II.1 nous avons vu qu'une mesure μ sur (X,d) satisfaisant $T_1(\lambda)$ vérifie l'inégalité

$$\sup_{[\varphi]_{lip} \leq 1} \mathbb{P}\left[\frac{1}{N}\sum_{i=1}^N \varphi(Y^i) - \int_X \varphi(x)\, d\mu(x) > \varepsilon\right] \leq e^{-\frac{\lambda}{2}N\varepsilon^2} \tag{24}$$

pour tous $N \geq 1$, $\varepsilon > 0$ et toutes variables aléatoires Y^i indépendantes et de loi μ, le coefficient λ apparaissant dans le membre de droite étant le meilleur possible. Comme nous l'avons annoncé au début de cette partie II, nous voulons maintenant avoir la majoration plus forte

$$\mathbb{P}\left[\sup_{[\varphi]_{lip} \leq 1} \left|\frac{1}{N}\sum_{i=1}^N \varphi(Y^i) - \int_X \varphi(x)\, d\mu(x)\right| > \varepsilon\right] \leq e^{-\frac{\lambda}{2}N\varepsilon^2}, \tag{25}$$

c'est-à-dire

$$\mathbb{P}\big[W_1(\hat{\mu}^N, \mu) > \varepsilon\big] \leq e^{-\frac{\lambda}{2}N\varepsilon^2} \tag{26}$$

où $\hat{\mu}^N$ est la mesure empirique $\displaystyle\frac{1}{N}\sum_{i=1}^N \delta_{Y^i}$ des Y^i.

Dans le chapitre 6 nous montrons en particulier dans le cas où X est l'espace euclidien $\mathbb{R}^d$:

Théorème 10 (cf. Theorem 6.1). *Soient $p \in [1,2]$ et μ une mesure de probabilité sur $\mathbb{R}^d$ vérifiant une inégalité $T_p(\lambda)$. Alors, pour tous $d' > d$ et $\lambda' < \lambda$, il existe une constante N_0, dépendant de λ', d' et d'un moment exponentiel carré de μ, telle que*

$$\mathbb{P}\left[W_p(\hat{\mu}^N, \mu) > \varepsilon\right] \leq e^{-\gamma_p \frac{\lambda'}{2} N \varepsilon^2} \tag{27}$$

pour tous $\varepsilon > 0$, $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$ et toutes variables aléatoires $Y^1, \ldots, Y^N$ indépendantes et de loi μ, où $\hat{\mu}^N$ est leur mesure empirique et

$$\gamma_p = \begin{cases} 1 & si \quad 1 \leq p < 2 \\ 3 - 2\sqrt{2} & si \quad p = 2. \end{cases}$$

Pour $p = 1$ nous pouvons donc passer de (24) à (25) (ou (26)) au prix du remplacement de λ par $\lambda' < \lambda$ arbitrairement proche de λ, et d'une condition sur la taille N de l'échantillon. En fait une variante de la preuve de ce théorème assure une estimation de la forme

$$\mathbb{P}\left[W_p(\hat{\mu}^N, \mu) > \varepsilon\right] \leq C(\varepsilon)\, e^{-\gamma_p \frac{\lambda'}{2} N \varepsilon^2}$$

sans restriction sur N, mais pour une (grande) constante $C(\varepsilon)$, calculable d'après la preuve.

Le résultat du théorème 10 semble naturel au vu du théorème de Sanov (donné dans [43] par exemple). En effet, une application de ce théorème à l'ensemble $A = \{\nu; W_p(\nu,\mu) > \varepsilon\}$, pour un $\varepsilon > 0$ fixé, laisse espérer une majoration de la forme

$$\mathbb{P}\,[W_p(\hat{\mu}^N,\mu) > \varepsilon] \leq \exp\Big(- N \inf\big\{H(\nu|\mu); \nu \in A\big\}\Big)$$

pour N grand. Cette majoration étant admise, comme

$$\inf\big\{H(\nu|\mu); \nu \in A\big\} \geq \frac{\lambda}{2}\,\varepsilon^2$$

puisque μ satisfait une inégalité $T_p(\lambda)$, on obtient bien une majoration du type (27), mais seulement de manière asymptotique, alors que le théorème 10 donne de plus une estimation précise sur une taille suffisante de l'échantillon. En réalité ce théorème de Sanov ne donne pas une telle majoration ; en effet, sur l'espace non borné $\mathbb{R}^d$, la fermeture $\overline{A}$ de A pour la topologie faible (étroite) contient la mesure μ elle-même : ainsi $\inf\{H(\nu|\mu);\, \nu \in \overline{A}\} = 0$ et le théorème de Sanov ne donne alors que la majoration triviale $\mathbb{P}\,[W_p(\hat{\mu}^N,\mu) > \varepsilon] \leq \exp\big(- N \inf\{H(\nu|\mu); \nu \in \overline{A}\}\big) = 1$.

L'idée de la preuve du théorème 10 est la suivante. Dans une première étape nous ramenons le problème au cas d'un ensemble compact en tronquant les différents termes en dehors d'une boule de $\mathbb{R}^d$. Puis nous recouvrons l'ensemble des mesures de probabilité sur cette boule, qui est lui même compact, par un nombre fini de boules en distance de Wasserstein, sur lesquelles nous développons l'argument de Sanov. Enfin nous optimisons les paramètres introduits au cours de ces deux étapes. La condition $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$ imposée sur la taille de l'échantillon provient en particulier d'un calcul d'entropie métrique effectué dans la deuxième étape. Nous ne savons pas si cette condition peut être améliorée.

Pour des mesures satisfaisant des hypothèses plus faibles, par exemple n'admettant qu'un moment polynomial fini, nous donnerons dans les théorèmes 6.5 et 6.6 des versions plus faibles de ce résultat, mais fondées sur la même démarche.

Ce théorème sera appliqué dans la partie III à des variables aléatoires indépendantes $Y_t^{i,N}$ pour $1 \leq i \leq N$, évoluant avec le temps t et définies à partir des positions $X_t^{i,N}$ de particules en interaction. Par la suite nous nous placerons au niveau des trajectoires elles-mêmes, et non plus au niveau des positions des particules en un instant t fixé. Pour cela nous serons amenés à considérer des mesures sur l'espace $\mathcal{C}$ des fonctions continues de $[0,T]$ dans $\mathbb{R}^d$, pour un $T \geq 0$ fixé, muni de la norme uniforme et à les comparer à l'aide de la distance de Wasserstein W_p définie sur $\mathcal{P}_p(\mathcal{C})$ à partir de la norme uniforme sur $\mathcal{C}$. Notant $\mathcal{C}^\alpha$ l'espace des fonctions de $\mathcal{C}$ höldériennes d'indice $\alpha \in]0,1]$, muni de la norme höldérienne

$$\|f\|_\alpha = \sup\Big(\sup_{0\leq t\leq T} |f(t)|, \sup_{\substack{0\leq t,s\leq T\\ t\neq s}} \frac{|f(t)-f(s)|}{|t-s|^\alpha}\Big),$$

et qui est un borélien de $\mathcal{C}$, nous montrons dans le chapitre 7 l'extension suivante du théorème 10 pour des mesures sur $\mathcal{C}$, concentrées sur $\mathcal{C}^\alpha$:

Théorème 11 (cf. Theorem 7.1). *Soient $p \in [1,2]$ et μ une mesure de probabilité sur $\mathcal{C}$ vérifiant une inégalité $T_p(\lambda)$ pour un $\lambda > 0$, et telle que $\int_{\mathcal{C}} e^{a\|x\|_\alpha^2}\, d\mu(x)$ soit fini pour un $a > 0$ et un $\alpha \in]0,1]$. Alors, pour tous $\alpha' < \alpha$ et $\lambda' < \lambda$, il existe une constante N_0 telle que*

$$\mathbb{P}\left[W_p(\mu, \hat{\mu}^N) > \varepsilon\right] \leq e^{-\beta_p \frac{\lambda'}{2} N \varepsilon^2} \tag{28}$$

pour tous $\varepsilon > 0$, $N \geq N_0\, \varepsilon^{-2} \exp\left(N_0\, \varepsilon^{-1/\alpha'}\right)$ et toutes variables aléatoires $Y^1, \ldots, Y^N$ indépendantes et de loi μ, où $\hat{\mu}^N$ est leur mesure empirique et

$$\beta_p = \begin{cases} 1 & si \quad 1 \leq p < 2 \\ (1 + \sqrt{\lambda/a})^{-2} & si \quad p = 2. \end{cases}$$

Cet énoncé est bien une extension du théorème 10 . En effet, si m est une mesure sur $\mathbb{R}^d$ satisfaisant $T_p(\lambda)$, alors la loi μ d'un processus constant sur $[0,T]$ et de loi initiale m satisfait les hypothèses du théorème 11 (pour tout $a < \lambda/2$). Par projection en $t = 0$, l'inégalité (28) se réduit alors à (27) avec la même constante γ_p, cependant sous une condition plus forte sur la taille de l'échantillon.

Par comparaison au théorème 10 nous avons ajouté la condition que μ soit concentrée sur un espace $\mathcal{C}^\alpha$ avec $\int_{\mathcal{C}} e^{a\|x\|_\alpha^2}\, d\mu(x)$ fini pour un $a > 0$. En adaptant la preuve du théorème précédent à ce cadre de dimension infinie, nous commençons par une troncature à un ensemble compact de $\mathcal{C}$; grâce à cette hypothèse de concentration nous pouvons prendre une boule de $\mathcal{C}^\alpha$ pour la norme höldérienne, qui est compacte dans $\mathcal{C}$ pour la topologie de la norme uniforme. Cette condition est par exemple vérifiée par la mesure de Wiener sur $\mathcal{C}$, les trajectoires du mouvement brownien étant en particulier höldériennes d'indice α pour tout $\alpha < 1/2$, et par extension le sera par la loi du processus particulier considéré dans la partie III.

Dans le paragraphe II.1 nous avons vu qu'une mesure μ sur un espace métrique complet et séparable (X,d) vérifie une inégalité de transport T_1 (ou de concentration gaussienne) dès que le moment $\int_X e^{a\, d(x_0,x)^2}\, d\mu(x)$ est fini pour un $a > 0$ et un $x_0 \in X$ (cf. théorème 9). Ce résultat est en fait une conséquence du résultat suivant que nous montrons dans le chapitre 4 :

Théorème 12 (cf. Theorem 4.1). *Soient X un espace mesurable, μ et ν deux mesures de probabilité sur X et φ une fonction mesurable positive sur X. Alors*

(i) $$\|\varphi\,(\mu - \nu)\|_{TV} \leq \left(\frac{3}{2} + \ln \int_X e^{2\varphi(x)}\, d\mu(x)\right) \left(\sqrt{H(\nu|\mu)} + \frac{1}{2} H(\nu|\mu)\right);$$

(ii) $$\|\varphi\,(\mu - \nu)\|_{TV} \leq \sqrt{2} \left(1 + \ln \int_X e^{\varphi(x)^2}\, d\mu(x)\right)^{1/2} \sqrt{H(\nu|\mu)}.$$

Pour $\varphi \equiv 1$ nous retrouvons l'inégalité de Csiszár-Kullback-Pinsker

$$\|\nu - \mu\|_{TV} \leq c\, \sqrt{H(\nu|\mu)}$$

avec la constante non optimale $c = 2$ au lieu de $\sqrt{2}$ (la constante optimale pouvant être retrouvée en réécrivant la preuve de ce théorème dans le cas particulier où $\varphi \equiv 1$).

Si $\displaystyle\int_X e^{ad(x_0,x)^{2p}}\, d\mu(x)$ est fini pour un $a > 0$ et un $x_0 \in X$, nous en déduisons que

$$W_p(\mu,\nu) \ \leq CH(\nu|\mu)^{\frac{1}{2p}}$$

pour toute mesure ν sur X, où

$$C = 2 \inf_{x_0 \in X,\, a>0} \left(\frac{1}{2a} \Big(1 + \ln \int_X e^{ad(x_0,x)^{2p}}\, d\mu(x) \Big) \right)^{\frac{1}{2p}}$$

est fini (cf. **Corollary 4.4**). En particulier, pour $p = 1$, nous retrouvons la condition suffisante du théorème 9.

La quantité $\|\varphi\,(\mu - \nu)\|_{TV}$ a depuis été interprétée dans [56, Proposition VI. 7] comme le coût de transport entre μ et ν pour la fonction de coût

$$c(x,y) = (\varphi(x) + \varphi(y))\mathbf{1}_{x \neq y}.$$

Des inégalités de la forme de celles obtenues dans le théorème 12 sont alors vues comme des exemples d'inégalités dites « norme-entropie », qui généralisent l'inégalité T_1, et de manière analogue sont équivalentes à certaines inégalités de déviation (cf. [57]).

II.3. Inégalités de Sobolev logarithmiques

Nous avons vu qu'une inégalité de concentration gaussienne $CG(\lambda)$ est équivalente à une inégalité de transport $T_1(\lambda)$, et est donc impliquée par une inégalité $T_p(\lambda)$ pour tout $p \geq 1$. Sur l'espace euclidien $\mathbb{R}^d$, elle est également impliquée par une *inégalité de Sobolev logarithmique* (ISL) ainsi définie : une mesure borélienne μ de probabilité sur l'espace euclidien $\mathbb{R}^d$ vérifie une ISL de constante λ, notée $ISL(\lambda)$, si

$$\int_{\mathbb{R}^d} f^2 \, \ln f^2(x)\, d\mu(x) - \int_{\mathbb{R}^d} f^2(x)\, d\mu(x)\, \ln \Big(\int_{\mathbb{R}^d} f^2(x)\, d\mu(x) \Big) \leq \frac{2}{\lambda} \int_{\mathbb{R}^d} |\nabla f|^2(x)\, d\mu(x)$$

pour toute fonction f de $\mathbb{R}^d$ dans $\mathbb{R}$ de classe $\mathcal{C}^\infty$ et à support compact.

Depuis que L. Gross [58] a montré que la mesure gaussienne vérifie cette inégalité avec $\lambda = 1$, de nombreux travaux ont porté et portent encore actuellement sur ces inégalités, tant du point de vue de leur étude théorique, avec l'établissement de conditions nécessaires ou suffisantes ou leur étude sur des espaces plus généraux, que de leurs liens avec d'autres domaines tels que les inégalités de Sobolev et de Poincaré, la théorie de l'information, l'hypercontractivité de semi-groupes, la concentration de la mesure, l'isopérimétrie, l'analyse de modèles issus de la mécanique statistique, ou enfin l'étude de la convergence vers l'équilibre de solutions de certaines équations aux dérivées partielles par des techniques de dissipation d'entropie. Tous ces aspects sont présentés dans [5], [67], [68, Chapter 5], [69], [111, Section 9.2] et dans les travaux mentionnés dans ces références.

Nous n'abordons ici que trois points, à savoir les liens entre ces inégalités et les inégalités de concentration et de transport, des conditions nécessaires ou suffisantes à leur obtention, et leurs propriétés de tensorisation.

Comme nous l'avons annoncé ci-dessus, une mesure sur $\mathbb{R}^d$ vérifie une inégalité $CG(\lambda)$ ou $T_1(\lambda)$ dès qu'elle vérifie $ISL(\lambda)$, comme l'assure l'argument de Herbst (cf. [5, Chapter 7] ou [68, Chapter 5] par exemple pour une démonstration et une discussion sur l'historique de ce résultat).

Il peut donc être intéressant de savoir quelles mesures vérifient une ISL. En dimension $d = 1$, ces mesures ont été caractérisées par S. Bobkov et F. Götze [17]. Sur $\mathbb{R}^d$ avec $d \geq 2$ il n'existe pas de telle caractérisation, mais des conditions suffisantes, de deux ordres (ainsi que des conditions nécessaires). Citons d'une part le critère suivant dû à D. Bakry et M. Emery [8] (cf. [16], [18], [31], [40], [69], [88] pour d'autres démonstrations, dont certaines sont fondées sur le transport optimal) : toute mesure sur $\mathbb{R}^d$ ayant une densité de la forme $e^{-V(x)}$ par rapport à la mesure de Lebesgue, avec V de classe $\mathcal{C}^2$ tel que $D^2V(x) \geq \lambda \, Id$ pour un $\lambda > 0$ et tout $x \in \mathbb{R}^d$, vérifie $ISL(\lambda)$; en particulier la mesure gaussienne sur $\mathbb{R}^d$ vérifie $ISL(1)$ et ceci quelle que soit la dimension d (nous reviendrons sur cet aspect indépendant de la dimension). Citons d'autre part les critères suivants qui permettent de transmettre une ISL d'une mesure à une autre :

• si une mesure μ sur $\mathbb{R}^d$ vérifie $ISL(\lambda)$ et v est une fonction bornée sur $\mathbb{R}^d$, alors $e^{-v}\,\mu$ vérifie l'inégalité $ISL\big(\lambda \exp[-2(\sup v - \inf v)]\big)$; pour de tels v nous avons noté dans le paragraphe II.1 que $e^{-v}\,\mu$ vérifie une inégalité T_1 s'il en est de même de μ ;

• si une mesure μ sur $\mathbb{R}^d$ vérifie $ISL(\lambda)$ et φ est une fonction lipschitzienne de $\mathbb{R}^d$ dans $\mathbb{R}^n$, de semi-norme de Lipschitz L, alors la mesure image de μ par φ vérifie $ISL(\lambda/L^2)$: prenant pour μ la mesure gaussienne, cette propriété permet par exemple de montrer que la loi uniforme sur $[0, 1]$ vérifie une ISL (cf. [54] pour la constante optimale dans cette inégalité) ou de retrouver le critère de Bakry - Emery à partir d'un résultat général de L. Caffarelli [31, Theorem 11], repris dans [111, Theorem 9.14] ; nous nous servirons également de cette propriété dans le paragraphe 5.3 du chapitre 5 ;

• si μ_1 et μ_2 sont deux mesures sur $\mathbb{R}^d$ et $\mathbb{R}^n$ respectivement vérifiant $ISL(\lambda)$, alors $\mu_1 \otimes \mu_2$ satisfait aussi $ISL(\lambda)$ sur l'espace euclidien $\mathbb{R}^d \times \mathbb{R}^n$: cette propriété, dite de tensorisation, est fondamentale ; elle assure en particulier que si μ vérifie $ISL(\lambda)$ sur $\mathbb{R}^d$, alors la mesure produit $\mu^{\otimes N}$ vérifie aussi $ISL(\lambda)$ sur $(\mathbb{R}^d)^N$ pour tout $N \geq 1$, avec donc une constante indépendante de N (ce qui par exemple n'est pas le cas des inégalités de Sobolev pour la mesure de Lebesgue).

Après avoir noté que $ISL(\lambda)$ implique $T_1(\lambda)$ sur $\mathbb{R}^d$, comparons à présent les inégalités $ISL(\lambda)$ et $T_2(\lambda)$. Ces deux inégalités présentent les points communs suivants : d'une part elles sont vérifiées par toute mesure de la forme $e^{-V(x)}\,dx$ où $D^2V(x) \geq \lambda \, Id$ pour tout $x \in \mathbb{R}^d$; d'autre part elles sont stables par tensorisation et impliquent une inégalité $T_1(\lambda)$ (ainsi qu'une inégalité de Poincaré). A la suite de plusieurs travaux sur le sujet, on peut maintenant conclure que $ISL(\lambda)$ implique $T_2(\lambda)$ (cf. [16] et [88]), mais qu'une inégalité T_2 n'implique pas une ISL en toute généralité (cf. P. Cattiaux et A. Guillin [39] pour un contre-exemple en dimension 1) ; cependant il existe des formes partielles de réciproque pour des mesures de la forme $e^{-V(x)}\,dx$ avec $D^2V(x) \geq \alpha\, Id$ pour un $\alpha \in \mathbb{R}$ et tout $x \in \mathbb{R}^d$ (cf. [40] et [88]).

Pour des mesures sur l'espace euclidien $\mathbb{R}^d$ nous avons donc le schéma

$$ISL(\lambda) \;\Rightarrow\; T_2(\lambda) \;\Rightarrow\; T_1(\lambda) \;\Leftrightarrow\; CG(\lambda)$$

où les deux implications sont strictes sans hypothèse supplémentaire.

Dans le prochain paragraphe nous nous intéresserons à ces inégalités de concentration gaussienne, de transport et de Sobolev logarithmique sur des espaces produits.

II.4. Le cas des espaces produits (cf. chapitres 4 et 5*)

Du paragraphe II.2 il ressort que si, pour $i = 1, \ldots, N$, la mesure μ_i sur l'espace métrique (X, d) vérifie une inégalité $T_2(\lambda)$, alors la mesure produit $\mu_1 \otimes \cdots \otimes \mu_N$ vérifie également une inégalité $T_2(\lambda)$ sur l'espace produit X^N muni de la distance d_{ℓ^2} définie par $d_{\ell^2}(x,y) = \Big(\sum\limits_{i=1}^{N} d(x_i,y_i)^2\Big)^{1/2}$. D'après le paragraphe II.3 il en est de même pour les inégalités de Sobolev logarithmiques si X est l'espace euclidien $\mathbb{R}^d$ (et donc X^N est l'espace euclidien $(\mathbb{R}^d)^N$).

La situation est différente pour l'inégalité T_1 ou de concentration gaussienne. Supposons en effet, avec les notations précédentes, que chaque μ_i vérifie $CG(\lambda)$: qu'en est-il alors de $\mu_1 \otimes \cdots \otimes \mu_N$ sur X^N ?

On peut tout d'abord montrer que $\mu_1 \otimes \cdots \otimes \mu_N$ vérifie $CG(\lambda/N)$ sur X^N muni de la distance d_{ℓ^2}. Interprétons ce résultat en termes d'inégalités de déviation pour des variables aléatoires. Soient par exemple $Y^1, \ldots, Y^N$ des variables indépendantes de même loi μ satisfaisant $CG(\lambda)$ sur X et soit φ une fonction lipschitzienne sur X de semi-norme de Lipschitz 1. L'application Φ définie sur X^N par $\Phi(x_1, \ldots, x_N) = \sum\limits_{i=1}^{N} \varphi(x_i)$ étant lipschitzienne de semi-norme $\sqrt{N}$ sur l'espace (X^N, d_{ℓ^2}) et $\mu^{\otimes N}$ vérifiant $CG(\lambda/N)$ sur cet espace, on a la majoration

$$\int_{X^N} e^{t\left(\frac{1}{\sqrt{N}}\Phi - \int_{X^N} \frac{1}{\sqrt{N}}\Phi\, d\mu^{\otimes N}\right)}\, d\mu^{\otimes N} \leq e^{\frac{Nt^2}{2\lambda}},$$

puis (seulement)

$$\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} \varphi(Y^i) - \int_X \varphi(x)\, d\mu(x) > \varepsilon\right] \leq e^{-\frac{\lambda}{2}\varepsilon^2} \tag{29}$$

pour tout $\varepsilon > 0$ d'après l'inégalité de Chebyschev. Or d'après le paragraphe II.1 on a en fait

$$\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} \varphi(Y^i) - \int_X \varphi(x)\, d\mu(x) > \varepsilon\right] \leq e^{-\frac{\lambda}{2}N\varepsilon^2}. \tag{30}$$

* Le chapitre 5 reprend l'article [15] écrit en collaboration avec G. Blower.

Cette propriété de tensorisation pour la distance d_{ℓ^2} ne semble donc pas intéressante. On peut l'améliorer en munissant X^N de la distance d_{ℓ^1} définie par $d_{\ell^1}(x,y) = \sum_{i=1}^{N} d(x_i, y_i)$, comme le suggère la définition de $CG(\lambda)$, en termes de fonctions lipschitziennes, ou de $T_1(\lambda)$, dans laquelle apparaît la distance d et non directement d^2 comme pour T_2. Il peut alors être montré que $\mu_1 \otimes \cdots \otimes \mu_N$ vérifie l'inégalité $CG(\lambda/N)$ et que la constante λ/N est la meilleure que l'on puisse obtenir. En termes d'inégalités de déviation, l'application Φ étant maintenant lipschitzienne de semi-norme 1 sur X^N puisque φ l'est sur X, l'inégalité $CG(\lambda/N)$ vérifiée par $\mu^{\otimes N}$ assure que

$$\int_{X^N} e^{t(\Phi - \int_{X^N} \Phi \, d\mu^{\otimes N})} \, d\mu^{\otimes N} \leq e^{\frac{Nt^2}{2\lambda}},$$

ce qui implique bien (30).

Nous voyons sur cet exemple que, bien que ces distances soient équivalentes sur $(\mathbb{R}^d)^N$ *pour N fixé* puisque $d_{\ell^2} \leq d_{\ell^1} \leq \sqrt{N} d_{\ell^2}$, la distance d_{ℓ^1} semble plus adaptée que d_{ℓ^2} à la tensorisation des inégalités de concentration gaussienne, puisqu'elle permet d'obtenir le bon ordre de convergence dans les inégalités de déviation.

D'après le théorème 9, une mesure μ sur X telle que $\int_X e^{ad(x_1,y_1)^2} \, d\mu(x_1)$ soit fini pour un $a > 0$ et un $y_1 \in X$, vérifie $T_1(\lambda)$ pour un certain λ. Dans ce cas, si X est muni de la distance d_{ℓ^2}, alors

$$\int_{X^N} e^{ad_{\ell^2}(x,y)^2} \, d\mu^{\otimes N}(x) = \Big(\int_X e^{ad(x_1,y_1)^2} \, d\mu(x_1) \Big)^N$$

est fini pour $y = (y_1, \ldots, y_1) \in X^N$, et donc $\mu^{\otimes N}$ vérifie l'inégalité de transport

$$W_1(\nu, \mu^{\otimes N}) \leq \left(\frac{2}{a} \Big(1 + N \ln \int_X e^{ad(x_1,y_1)^2} \, d\mu(x_1) \Big) \right)^{\frac{1}{2}} \sqrt{H(\nu|\mu^{\otimes N})}$$

pour toute mesure ν sur X^N, où W_1 est définie sur $\mathcal{P}(X^N)$ à partir de la distance d_{ℓ^2} sur X^N. Or, par exemple pour $\nu = \rho^{\otimes N}$ où ρ est une mesure sur X absolument continue par rapport à μ, le membre de gauche de cette inégalité de transport est d'ordre $\sqrt{N}$, alors que le membre de droite est d'ordre $\sqrt{N}.\sqrt{N} = N$.

Si maintenant X^N est muni de la distance d_{ℓ^1}, alors

$$\int_{X^N} e^{\frac{a}{N} d_{\ell^1}(x,y)^2} \, d\mu^{\otimes N}(x) \leq \Big(\int_X e^{ad(x_1,y_1)^2} \, d\mu(x_1) \Big)^N$$

est fini pour $y = (y_1, \ldots, y_1) \in X^N$ et donc $\mu^{\otimes N}$ vérifie l'inégalité de transport

$$W_1(\nu, \mu^{\otimes N}) \leq \left(\frac{2N}{a} \Big(1 + N \ln \int_{X^N} e^{ad(x_1,y_1)^2} \, d\mu(x_1) \Big) \right)^{\frac{1}{2}} \sqrt{H(\nu|\mu^{\otimes N})}$$

pour toute mesure ν sur X^N, où maintenant W_1 est définie sur $\mathcal{P}(X^N)$ à partir de la distance d_{ℓ^2} sur X^N. Dans le même exemple où $\nu = \rho^{\otimes N}$, le membre de gauche est d'ordre N alors que le membre de droite est d'ordre $N^{3/2}$.

Il semble donc que la caractérisation de l'inégalité de concentration gaussienne en termes de moments exponentiels carrés ne donne rien d'intéressant pour des mesures sur des espaces produits (aussi bien pour la distance d_{ℓ^1} que la distance d_{ℓ^2}). Il conviendra donc de limiter son utilisation à des mesures sur des espaces dont la dimension est fixée ou n'a pas d'impact particulier : ce sera le cas des mesures μ_t décrivant l'état macroscopique du système que l'on considérera dans la partie III de cette introduction, et qui sont des mesures sur l'espace des phases $\mathbb{R}^d$ (de dimension indépendante du nombre N de particules en interaction dans le modèle microscopique associé).

Nous venons donc de voir que si $Y^1, \ldots, Y^N$ sont N variables indépendantes de loi respective μ_i sur un espace (X, d) vérifiant une inégalité $CG(\lambda)$ (resp. $T_2(\lambda)$), alors la loi $\mu_1 \otimes \cdots \otimes \mu_N$ de $(Y^1, \ldots, Y^N)$ satisfait une inégalité $CG(\lambda/N)$ (resp. $T_2(\lambda)$) sur l'espace X^N muni de la distance d_{ℓ^1} (resp. d_{ℓ^2}). Si de plus X est l'espace euclidien $\mathbb{R}^d$ et si les μ_i vérifient une $ISL(\lambda)$, alors $\mu_1 \otimes \cdots \otimes \mu_N$ satisfait aussi $ISL(\lambda)$ sur l'espace euclidien $(\mathbb{R}^d)^N$.

Nous nous proposons maintenant de considérer certains cas de variables dépendantes. Plus précisément nous considérons un processus stochastique $(Y^1, \ldots, Y^N)$ à valeurs dans X et donnons des conditions suffisantes sur la distribution initiale $P^{(1)}$ de Y^1 et les distributions conditionnelles $p_i(. \mid x_1, \ldots, x_{i-1})$ de Y^i sachant $Y^1, \ldots, Y^{i-1}$, pour $2 \leq i \leq N$, pour que la loi

$$P^{(N)}(dx_1, \ldots, dx_N) = p_N(dx_n \mid x_1, \ldots, x_{N-1}) \ldots p_2(dx_2 \mid x_1)\, P^{(1)}(dx_1)$$

de $(Y^1, \ldots, Y^N)$ vérifie une inégalité de concentration gaussienne, de transport ou de Sobolev logarithmique.

Dans le cas de variables indépendantes, il suffit de supposer que $P^{(1)}$ (qui est μ_1 avec les notations précédentes) et que chaque $p_i(. \mid x_1, \ldots, x_{i-1})$ (qui est μ_i) vérifient une telle inégalité. Dans le cas de variables dépendantes nous ajoutons une condition sur les distributions conditionnelles, traduisant précisément le fait que chaque Y^i ne dépend pas trop de $Y^1, \ldots, Y^{i-1}$, c'est-à-dire que chaque $p_i(. \mid x_1, \ldots, x_{i-1})$ ne dépend pas trop de $x_1, \ldots, x_{i-1}$.

Traitant de questions analogues dans divers cadres, K. Marton [78, 79] et P.-M. Samson [96] traduisent cette condition, par exemple dans le cas de processus de Markov, par un coefficient α_i (nul dans le cas indépendant) tel que

$$\sup_{x_{i-1}, y_{i-1} \in X} D\big(p_i(. \mid x_{i-1}), p_i(. \mid y_{i-1})\big) \leq \alpha_i$$

où D mesure l'écart entre mesures de probabilité, comme la norme de variation totale de leur différence (cf. [96]).

Ici, suivant les travaux de H. Djellout, A. Guillin et L. Wu [46], nous considérons plutôt des conditions de la forme

$$\sup_{x_{i-1}, y_{i-1} \in X} \frac{D\big(p_i(. \mid x_{i-1}), p_i(. \mid y_{i-1})\big)}{d(x_{i-1}, y_{i-1})} \leq \alpha_i.$$

Dans le chapitre 5 nous donnons tout d'abord le résultat suivant de concentration gaussienne pour la loi jointe des N variables $Y^1, \ldots, Y^N$ d'un processus stochastique :

Théorème 13 (cf. Theorem 5.6). *Soit* $(Y^1, \ldots, Y^N)$ *un processus stochastique à valeurs dans* (X, d)*, de loi initiale* $P^{(1)}$ *et de distributions conditionnelles* $p_i(. \mid x_i, \ldots, x_{i-1})$*. Supposons qu'il existe* $\lambda > 0$*,* $L \geq 0$ *et* $\rho_1, \ldots, \rho_{N-1} \geq 0$ *avec* $\displaystyle\sum_{j=1}^{N-1} \rho_j \leq L$ *tels que*

(i) $P^{(1)}$ *et* $p_i(. \mid x_1, \ldots, x_{i-1})$ *pour* $x_j \in X, 1 \leq j \leq i-1$ *et* $2 \leq i \leq N$ *satisfassent* $CG(\lambda)$ *sur* X *;*

(ii) pour $x_j, y_j \in X, 1 \leq j \leq i-1$ *et* $2 \leq i \leq N$

$$W_1\big(p_i(. \mid x_1, \ldots, x_{i-1}), p_i(. \mid y_1, \ldots, y_{i-1})\big) \leq \sum_{j=1}^{i-1} \rho_{i-j}\, d(x_j, y_j).$$

Alors la loi jointe de $(Y^1, \ldots Y^N)$ *vérifie* $CG(\lambda_N)$ *sur* (X^N, d_{ℓ^1})*, avec*

$$\frac{1}{\lambda_N} = \frac{1}{\lambda} \sum_{m=1}^{N} \Big(\sum_{i=0}^{m-1} L^i\Big)^2.$$

La dépendance de Y^i en Y^j avec $j < i$ est ici mesurée par le coefficient ρ_{i-j}, qui ne dépend que de la différence $i - j$ entre les temps j et i, et est nul pour tout $i - j > 1$ dans le cas d'un processus de Markov et pour tout $i - j$ dans le cas de variables indépendantes.

Pour $L < 1$ nous retrouvons l'inégalité $CG(\lambda(1 - L)^2/N)$ de [46], qui étend à des variables faiblement dépendantes l'inégalité $CG(\lambda/N)$ vérifiée par la loi jointe de N variables indépendantes $Y^1, \ldots, Y^N$ (pour lesquelles $L = 0$) dont les lois respectives vérifient toutes $CG(\lambda)$. Ce résultat implique les inégalités de déviation

$$\sup_{[\varphi]_{lip} \leq 1} \mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} \varphi(Y^i) - \mathbb{E}\Big(\frac{1}{N}\sum_{i=1}^{N} \varphi(Y^i)\Big) > \varepsilon\right] \leq e^{-\frac{\lambda}{2}(1-L)^2 N \varepsilon^2} \tag{31}$$

pour tous $\varepsilon > 0$ et $N \geq 1$, et

$$\sup_{[\varphi]_{lip} \leq 1} \mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} \varphi(Y^i) - \int_{\mathbb{R}^d} \varphi(x)\, d\mu(x) > \varepsilon\right] \leq e^{-\frac{\lambda}{2}(1-L)^2 N \varepsilon^2}$$

si de plus les Y^i ont même loi μ.

Pour chaque valeur de L nous donnons dans le chapitre 5 (cf. **Example 5.15**) un exemple de processus de Markov $(Y^1, \ldots, Y^N)$, construit à partir du processus d'Ornstein-Uhlenbeck, qui satisfait les hypothèses du théorème 13 et dont la loi jointe $P^{(N)}$ vérifie une inégalité de concentration de constante *égale* à la constante λ_N donnée par le théorème : pour cela nous montrons que l'inégalité

$$\int_{X^N} e^{t(\varphi - \int_{X^N} \varphi\, dP^{(N)})}\, dP^{(N)} \leq e^{\frac{t^2}{2\lambda_N}}$$

est une égalité pour une certaine fonction lipschitzienne φ et tout réel t. En ce sens les constantes données par le théorème 13 sont optimales.

Dans le chapitre 4 (cf. **Theorem 4.13**) nous utilisons la caractérisation des inégalités de concentration gaussienne en termes de moments exponentiels carrés pour établir de telles inégalités pour la loi jointe d'un processus stochastique : les conditions de faible dépendance portent sur des moments exponentiels des mesures de transition, qui sont moins contraignantes que la condition (ii) du théorème 13 avec $L < 1$ (cf. **Example 4.17**), mais suffisantes pour établir une inégalité $CG(\lambda'/N)$ pour la loi jointe des N premières variables, où $\lambda' > 0$ est indépendant de N. Cependant ces résultats sont obtenus sur l'espace produit X^N muni de la distance d_{ℓ^2} et donc, comme nous l'avons déjà observé pour des variables indépendantes (cf. (29)), ne peuvent induire des inégalités de déviation de la forme (31).

Le théorème 13 est obtenu dans le chapitre 5 comme cas particulier (correspondant à $p = 1$) d'un résultat plus général (cf. **Theorem 5.4**) assurant une inégalité de transport T_p pour la loi jointe de processus stochastiques. Nous ne détaillons pas ce résultat dans cette introduction, et présentons maintenant un énoncé qui est un analogue du théorème 13 au niveau des inégalités de Sobolev logarithmiques.

Nous considérons désormais un processus stochastique $(Y^1, \dots, Y^N)$ à valeurs dans l'espace euclidien $\mathbb{R}^d$ et cherchons des conditions suffisantes sur les distributions conditionnelles pour que la loi jointe de $(Y^1, \dots, Y^N)$ vérifie une ISL sur $(\mathbb{R}^d)^N$, si possible avec une constante indépendante du nombre N de variables (ce qui étendrait le cas de variables indépendantes abordé au début de ce paragraphe).

Pour cela nous supposons que les distributions conditionnelles ont une densité strictement positive par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ et sont de la forme

$$p_i(dx_i \mid x_1, \dots, x_{i-1}) = e^{-u_i(x_1,\dots,x_i)} dx_i, \qquad 2 \le i \le N \tag{32}$$

pour des fonctions u_i de $(\mathbb{R}^d)^i$ dans $\mathbb{R}$. La dépendance de Y^i en Y^j pour $j < i$ se traduisant par la dépendance de la fonction $u_i = u_i(x_1, \dots, x_i)$ en la variable x_j, nous la mesurons par la transformée de Laplace

$$\Lambda_{i,j}(s) = \sup_{x_1,\dots,x_{i-1}} \int_{\mathbb{R}^d} \exp\Big(s \cdot \frac{\partial u_i}{\partial x_j}(x_1, \dots, x_i)\Big)\, p_i(dx_i \mid x_1, \dots, x_{i-1}), \qquad s \in \mathbb{R}^d,\ j < i$$

où $\partial/\partial x_j$ est le gradient par rapport à $x_j \in \mathbb{R}^d$ et $a \cdot b$ est le produit scalaire de $a, b \in \mathbb{R}^d$.

Dans le chapitre 5 nous montrons alors le

Théorème 14 (cf. Theorem 5.10). *Avec les notations introduites précédemment, soit $(Y^1, \dots, Y^N)$ un processus stochastique à valeurs dans $\mathbb{R}^d$, de loi initiale $P^{(1)}$ et de distributions conditionnelles $p_i(. \mid x_1, \dots, x_{i-1})$ pour $2 \le i \le N$ de la forme (32). Supposons qu'il existe des constantes $\lambda > 0$, $L \ge 0$ et $\rho_1, \dots, \rho_{N-1}$ avec $\displaystyle\sum_{\ell=1}^{N-1} \sqrt{\rho_\ell} \le \sqrt{L}$ telles que*

(i) $P^{(1)}$ et $p_i(. \mid x_1, \dots, x_{i-1})$ pour $x_j \in X, 1 \le j \le i-1$ et $2 \le i \le N$ vérifient $ISL(\lambda)$ sur $\mathbb{R}^d$;

(ii) $\Lambda_{i,j}(s) \le \exp(\rho_{i-j}\, |s|^2/2)$ pour tout $s \in \mathbb{R}^d$.

Alors la loi jointe de $(Y^1,\dots,Y^N)$ *vérifie* $ISL(\lambda_N)$ *sur* $(\mathbb{R}^d)^N$ *où*

$$\lambda_N = \begin{cases} \big(\sqrt{\lambda}-\sqrt{L}\big)^2 & si \quad L<\lambda, \\ \dfrac{\lambda}{N(N+1)(e-1)} & si \quad L=\lambda, \\ \Big(\dfrac{\lambda}{L}\Big)^N \dfrac{L-\lambda}{e(N+1)} & si \quad L>\lambda. \end{cases}$$

La dépendance de Y^i en Y^j avec $j<i$ est contrôlée par l'hypothèse (ii). Sous l'hypothèse (i), nous montrons dans le chapitre 5 (cf. **Proposition 5.11**) que cette hypothèse (ii) est satisfaite pour $j<i$ donnés dès que u_i est deux fois différentiable sur $(\mathbb{R}^d)^i$, avec

$$-\sqrt{\lambda\,\rho_{i-j}}\,Id \;\le\; \frac{\partial^2 u_i}{\partial x_i \partial x_j}(x_1,\dots,x_i) \;\le\; \sqrt{\lambda\,\rho_{i-j}}\,Id, \qquad x_1,\dots,x_i\in\mathbb{R}^d$$

au sens des matrices symétriques de taille $d\times d$.

Par exemple pour un processus de Markov nous en déduisons le

Corollaire 15 (cf. Theorem 5.3). *Soit* $(Y^1,\dots,Y^N)$ *un processus de Markov à valeurs dans* $\mathbb{R}^d$*, de loi initiale* $P^{(1)}$ *et de noyau de transition de la forme* $p(dy \mid x) = e^{-u(x,y)}\,dy$*. Supposons qu'il existe des constantes* $\lambda>0$ *et* $L\ge 0$ *telles que*

(i) $P^{(1)}$ *et* $p(.\mid x)$ *pour* $x\in\mathbb{R}^d$ *vérifient* $ISL(\lambda)$ *dans* $\mathbb{R}^d$ *;*

(ii) u *soit deux fois différentiable sur* $\mathbb{R}^d\times\mathbb{R}^d$ *et vérifie*

$$-L\,Id \;\le\; \frac{\partial^2 u}{\partial x\partial y}(x,y) \;\le\; L\,Id, \qquad x,y\in\mathbb{R}^d$$

au sens des matrices symétriques de taille $d\times d$*.*

Alors la loi jointe de $(Y^1,\dots,Y^N)$ *vérifie* $ISL(\lambda_N)$ *sur* $(\mathbb{R}^d)^N$ *où*

$$\lambda_N = \begin{cases} \dfrac{(\lambda-L)^2}{\lambda} & si \quad L<\lambda, \\ \dfrac{\lambda}{N(N+1)(e-1)} & si \quad L=\lambda, \\ \Big(\dfrac{\lambda}{L}\Big)^{2N} \dfrac{L^2-\lambda^2}{\lambda e(N+1)} & si \quad L>\lambda. \end{cases}$$

En particulier la loi jointe de $(Y^1,\dots,Y^N)$ vérifie une ISL de constante uniforme en N si la dépendance entre les variables n'est pas trop grande au sens où $L<\lambda$; ceci étend l'inégalité $ISL(\lambda)$ dans le cas de variables indépendantes pour lequel $L=0$.

Comme nous l'avons fait pour les inégalités de concentration gaussienne, nous montrons dans le chapitre 5 (cf. **Example 5.15**) que pour chaque valeur de L l'ordre de grandeur en N de la constante λ_N obtenue dans le corollaire 15 est optimal sans hypothèse supplémentaire.

Ce problème d'obtention d'une ISL pour la loi jointe de variables dépendantes a été considéré indépendamment et dans un cadre différent par K. Marton [80]. Les résultats obtenus peuvent être comparés dans la situation simple du corollaire 15 : cet énoncé assure une ISL pour la loi jointe, avec une constante indépendante du nombre N de variables, dès que $L < \lambda$; le résultat principal de [80] (cf. Theorem), dans ce cas particulier, semble n'assurer un tel résultat que pour $L < \lambda/2$.

Notons enfin que, toute mesure P sur l'espace produit $(\mathbb{R}^d)^N$ pouvant se décomposer en une distribution « initiale » $P^{(1)}$ et $N-1$ distributions conditionnelles $p_i(. \mid x_1, \ldots, x_{i-1})$ pour $2 \leq i \leq N$ sous la forme

$$P(dx_1, \ldots, dx_N) = p_N(dx_n \mid x_1, \ldots, x_{N-1}) \ldots p_2(dx_2 \mid x_1)\, P^{(1)}(dx_1)$$

et rentrant donc potentiellement dans le cadre du théorème 14, celui-ci ne permet pas de retrouver toutes les ISL connues pour des mesures sur des espaces produits. Ainsi, des mesures intervenant dans des modèles de mécanique statistique, considérés dans [59] et [69] par exemple, vérifient des ISL avec des constantes indépendantes de la dimension N, ce qui ne semble pas se déduire directement du théorème 14.

Ce théorème 14 n'est donc qu'un résultat partiel vers la caractérisation des mesures sur les espaces produits vérifiant une ISL, si possible avec une constante indépendante de la dimension. Il serait en particulier intéressant d'établir une forme de réciproque de cet énoncé pour déterminer dans quelle mesure il approche une telle caractérisation, et d'affaiblir les hypothèses afin d'inclure des exemples tels que ceux considérés dans [59] et [69].

III - Limites de champ moyen pour des systèmes de particules stochastiques en interaction

Dans la partie I de cette introduction nous avons abordé l'étude de limites de champ moyen par l'approximation des équations de Vlasov et d'Euler par des systèmes de particules déterministes en interaction. Nous nous tournons maintenant vers un problème analogue pour des équations aux dérivées partielles comportant un terme de diffusion, ce que nous traduirons au niveau de l'approximation par une évolution aléatoire des particules.

III.1. Présentation du modèle macroscopique

Nous nous proposons d'étudier une approximation particulaire de certaines équations aux dérivées partielles d'évolution, de la forme

$$\frac{\partial \mu}{\partial t} = \Delta \mu + \nabla \cdot (\mu \nabla V) + \nabla \cdot (\mu (\nabla W * \mu)), \qquad t > 0, \quad x \in \mathbb{R}^d \tag{33}$$

où Δ est le laplacien, ∇ le gradient et $\nabla \cdot$ l'opérateur de divergence sur $\mathbb{R}^d$. Nous considérerons des solutions au sens des distributions, avec des mesures de probabilité sur $\mathbb{R}^d$ comme données initiales. Comme l'équation préserve la positivité et la masse totale, les solutions seront à chaque instant des mesures de probabilité sur $\mathbb{R}^d$. D'après la formule d'Itô, la valeur μ_t à l'instant t d'une telle solution μ de donnée initiale μ_0 peut être vue comme la distribution d'une variable aléatoire Y_t à valeurs dans $\mathbb{R}^d$, évoluant suivant l'équation différentielle stochastique

$$dY_t = \sqrt{2}\,dB_t - \nabla V(Y_t) - \nabla W * \mu_t(Y_t)$$

à partir d'une donnée initiale Y_0 distribuée suivant μ_0, pour un mouvement brownien $(B_t)_{t\geq 0}$ à valeurs dans $\mathbb{R}^d$. En particulier l'évolution de $\mu_t(x) = \mu(t,x)$ en un point $x \in \mathbb{R}^d$ fixé (resp. l'évolution de Y_t) dépend de la valeur de $\mu_t(y)$ en tous les points $y \in \mathbb{R}^d$, à travers le terme de convolution $\nabla W * \mu_t(x)$ (resp. $\nabla W * \mu_t(Y_t)$) : en ce sens l'équation (33) est dite *de champ moyen*, comme le sont les équations de Vlasov et d'Euler considérées précédemment. Dans l'interprétation particulaire donnée par le processus $(Y_t)_{t\geq 0}$, les potentiels V et W doivent être respectivement interprétés comme des potentiels extérieur et d'interaction entre les différentes parties du système ; en particulier nous supposerons que W est une fonction paire sur $\mathbb{R}^d$.

En dimension $d = 1$ une équation de cette forme, pour les potentiels $V \equiv 0$ et $W(x) = |x|^3/3$, régit l'évolution de la densité macroscopique (en la variable de vitesse) d'un milieu granulaire homogène en espace et composé, au niveau microscopique, de particules interagissant par des chocs inélastiques et excitées par une source de chaleur (cf. [11], [12]).

Les équations (33) sont des exemples d'équations de McKean-Vlasov, de la forme générale

$$\frac{\partial \mu}{\partial t} = \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}(a_{ij}[x,\mu]\mu) + \sum_{i=1}^{d} \frac{\partial}{\partial x_i}(b_i[x,\mu]\mu), \qquad t > 0, \quad x \in \mathbb{R}^d$$

d'inconnue $\mu = \mu(t,x)$. Ici les $a_{ij}[x,\mu]$ sont les coefficients de la matrice $a[x,\mu] = s[x,\mu]^* s[x,\mu]$ avec $s[x,\mu] = \displaystyle\int_{\mathbb{R}^d} \sigma(x,y)\,\mu(y)\,dy$ pour une famille $\sigma(x,y)$ de matrices $d \times d$ et les $b_i[x,v]$ sont les composantes du vecteur $\displaystyle\int_{\mathbb{R}^d} \beta(x,y)\mu(y)\,dy$ pour une famille $\beta(x,y)$ de vecteurs de $\mathbb{R}^d$ (cf. [73]). A cette famille d'équations appartient en particulier l'équation incompressible de Navier-Stokes dans le plan, en formulation tourbillon, de la forme

$$\frac{\partial \omega}{\partial t} = \nu\,\Delta\omega - \nabla \cdot (\omega\, u), \qquad t > 0, \quad x \in \mathbb{R}^2$$

où $u_t = K * \omega_t$ est le champ de vitesse du fluide au temps t, $K = \Big(\dfrac{\partial G}{\partial x_2}, -\dfrac{\partial G}{\partial x_1}\Big)$ et $G(x) = -\dfrac{1}{2\,\pi} \ln|x|$ est la solution fondamentale de l'équation de Poisson sur $\mathbb{R}^2$. Cette équation s'obtient à partir des équations incompressibles de Navier-Stokes dans le plan de la même manière que l'équation d'Euler (8) sur le tourbillon se déduit des équations d'Euler (7) sur la vitesse dans le cas non-visqueux où $\nu = 0$.

L'équation (33) a récemment été étudiée pour son comportement asymptotique en temps grand. Des résultats de convergence des solutions vers une solution stationnaire ont été obtenus par différentes techniques sous des hypothèses de convexité sur les potentiels V et W : nous aborderons ce point plus en détail dans le paragraphe III.6.

III.2. Approximation particulaire de l'équation

Dans le paragraphe précédent nous avons interprété la solution μ_t au temps t comme la distribution de Y_t évoluant suivant

$$dY_t = \sqrt{2}\, dB_t - \nabla V(Y_t) - \nabla W * \mu_t(Y_t). \tag{34}$$

Cette équation est parfois dite non-linéaire dans le sens où Y_t évolue suivant un champ de force $\nabla W * \mu_t$ dépendant de sa loi μ_t.

L'approximation particulaire de l'équation (33) consiste en l'introduction de N processus $(X_t^{i,N})_{t\geq 0}$ pour $1 \leq i \leq N$, représentant des particules évoluant chacune dans un champ de force créé non plus par la distribution μ_t, mais par les $N-1$ autres particules.

Nous avons vu dans la première partie de cette introduction que l'état d'un système constitué d'un grand nombre N de particules déterministes X^i, pour $1 \leq i \leq N$, dans l'espace des phases $\mathbb{R}^d$ peut être décrit par sa mesure empirique $\hat{m}^N = \dfrac{1}{N}\displaystyle\sum_{i=1}^N \delta_{X^i}$ puisque celle-ci génère les observables du système. Cette mesure empirique est en fait une densité de particules dans $\mathbb{R}^d$ dans la mesure où, pour tout borélien A de $\mathbb{R}^d$, la quantité $\hat{m}^N[A]$ est égale à la portion de particules parmi $X^1, \dots, X^N$ se trouvant dans la configuration A.

Supposons maintenant que ces particules X^i soient des variables aléatoires, échangeables dans le sens où leur loi jointe $m^{(N)}$ sur $(\mathbb{R}^d)^N$ est invariante par permutation de ses facteurs. Alors la loi d'une particule (quelconque par symétrie de la loi), qui est la première marginale de $m^{(N)}$ et donc une mesure déterministe sur $\mathbb{R}^d$, est égale à l'espérance $\mathbb{E}\,\hat{m}^N$ de la mesure empirique. Dans ce cas une notion de densité de particules est donc donnée par l'espérance de la mesure empirique.

Si $\hat{\mu}_t^N = \dfrac{1}{N}\displaystyle\sum_{i=1}^N X_t^{i,N}$ est la mesure empirique des N particules $X_t^{i,N}$ destinées à donner une approximation de l'équation (33), il peut donc sembler raisonnable de faire évoluer les $X_t^{i,N}$ dans le champ de force créé par la mesure $\hat{\mu}_t^N$, selon les équations différentielles stochastiques

$$dX_t^{i,N} = \sqrt{2}\, dB_t^{i,N} - \nabla V(X_t^{i,N}) - \nabla W * \hat{\mu}_t^N(X_t^{i,N}), \qquad 1 \leq i \leq N$$

c'est-à-dire

$$dX_t^{i,N} = \sqrt{2}\, dB_t^{i,N} - \nabla V(X_t^{i,N}) - \frac{1}{N}\sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N}), \qquad 1 \leq i \leq N \tag{35}$$

où les $(B_t^{i,N})_{1\leq i\leq N}$ sont N mouvements browniens indépendants. Supposant que les positions initiales $X_0^{i,N}$ soient échangeables, la symétrie du système d'équations assure qu'il en est de même pour les positions $X_t^{i,N}$ à tout instant ultérieur.

Dans cette situation d'évolution stochastique la mesure empirique $\hat{\mu}^N : t \mapsto \hat{\mu}_t^N$ ne vérifie pas l'équation originale (33), comme c'était le cas dans le cadre déterministe des équations de Vlasov ou d'Euler.

On peut cependant se demander si son espérance $\mathbb{E}\,\hat{\mu}^N$, qui à tout instant t associe la loi $\mu_t^{(1)}$ d'une particule quelconque du système, vérifie une telle équation. Dans cette optique, la formule d'Itô assure que $\mu^{(N)} : t \mapsto \mu_t^{(N)}$, où $\mu_t^{(N)}$ est la loi jointe du N-uplet $(X_t^{1,N}, \dots, X_t^{N,N})$, est solution de l'équation

$$\frac{\partial \mu^{(N)}}{\partial t} = \Delta^{(N)} \mu^{(N)} + \nabla^{(N)} \cdot \big(\mu^{(N)} \mathcal{V}\big) + \nabla^{(N)} \cdot \big(\mu^{(N)} (\mathcal{W} * \mu^{(N)})\big), \qquad t > 0, \quad x \in (\mathbb{R}^d)^N \quad (36)$$

où $\Delta^{(N)}$ et $\nabla^{(N)}\cdot$ sont le laplacien et l'opérateur de divergence sur $(\mathbb{R}^d)^N$,

$$\mathcal{V}(x) = \big(\nabla V(x^1), \dots, \nabla V(x^N)\big)$$

et

$$\mathcal{W}(x) = \Big(\frac{1}{N} \sum_{j=1}^N \nabla W(x^1 - x^j), \cdots, \frac{1}{N} \sum_{j=1}^N \nabla W(x^N - x^j)\Big)$$

pour $x = (x^1, \dots, x^N) \in (\mathbb{R}^d)^N$, et où ∇ est comme précédemment le gradient défini sur $\mathbb{R}^d$. Notant $\mu_t^{(2)}$ la loi jointe de deux particules du système au temps t et projetant l'équation (36) sur la première coordonnée, il s'ensuit que $\mu^{(1)}$ vérifie l'équation

$$\frac{\partial \mu^{(1)}}{\partial t} = \Delta \mu^{(1)} + \nabla \cdot (\mu^{(1)} \nabla V) + \nabla \cdot \Big(\int_{\mathbb{R}^d} \nabla W(x-y) \mu^{(2)}(x,y)\, dy\Big), \qquad t > 0, \quad x \in \mathbb{R}^d$$

au sens des distributions. Cette équation n'est pas fermée puisque, les particules $X_t^{i,N}$ n'étant pas indépendantes du fait de leur interaction (cf. (35)), la loi jointe $\mu_t^{(2)}$ de deux particules n'est pas égale au produit tensoriel $\mu_t^{(1)} \otimes \mu_t^{(1)}$ de la loi de chaque particule. Cependant, si tel était le cas, la mesure $\mu^{(1)}$ vérifierait l'équation (33), avec de plus la même donnée initiale μ_0 si les X_0^i ont pour distribution initiale μ_0 ; sous un résultat d'unicité pour (33), les mesures $\mu_t^{(1)}$ et μ_t seraient alors égales pour tout t.

L'interaction entre deux particules données semblant être d'ordre $1/N$ d'après l'équation (35), nous pouvons espérer que pour N grand elles se comportent de manière indépendante l'une de l'autre : nous reviendrons dans un instant sur ce phénomène, appelé *propagation du chaos* (cf. [105]) ou *propriété de Boltzmann* (cf. [63]). D'après ce qui précède nous pouvons alors espérer que pour N grand les mesures $\mu_t^{(1)}$, c'est-à-dire $\mathbb{E}\,\hat{\mu}_t^N$, et μ_t restent proches pour tout t si elles le sont initialement pour $t = 0$.

III.3. Premiers résultats de convergence

Ici nous voulons savoir si la mesure empirique aléatoire $\hat{\mu}_t^N$ des N particules $X_t^{i,N}$, et non seulement son espérance $\mathbb{E}\,\hat{\mu}_t^N$, est proche de la mesure déterministe μ_t, solution de (33) au temps t.

Cette question a été étudiée en particulier dans des travaux de H. McKean, H. Tanaka [109], A.-S. Sznitman [105], S. Méléard [84] ou S. Benachour, B. Roynette, D. Talay et P. Vallois [9, 10] (mais aussi pour d'autres équations, comme l'équation des milieux poreux dans [91]) en liaison avec le phénomène de propagation du chaos : sous diverses hypothèses sur la dimension d et les potentiels V et W, si les données initiales $X_0^{i,N}$ sont des variables indépendantes de loi μ_0, alors pour chaque k fixé la loi $\mu_t^{(k)}$ du k-uplet $(X_t^{1,N}, \ldots, X_t^{k,N})$ converge au sens de la topologie faible (étroite) des mesures sur $(\mathbb{R}^d)^k$ vers $\mu_t^{\otimes k}$ quand N tend vers l'infini : autrement dit les k variables $X_t^{1,N}, \ldots, X_t^{k,N}$ se comportent, asymptotiquement quand N tend vers l'infini, comme des variables indépendantes de loi μ_t. En particulier, pour $k = 1$, la loi $\mu_t^{(1)} (= \mathbb{E}\, \hat{\mu}_t^N)$ d'une particule converge bien vers μ_t.

D'après [105, Proposition 2.2] ou [84, Proposition 4.2], cette propriété est équivalente à la convergence en loi de la mesure empirique aléatoire $\hat{\mu}_t^N$ vers la mesure déterministe μ_t (comme éléments de l'espace des mesures de probabilité muni de la topologie faible (étroite)), et assure en particulier que

$$\mathbb{E}\left| \int_{\mathbb{R}^d} \varphi(x)\, d\hat{\mu}_t^N(x) - \int_{\mathbb{R}^d} \varphi(x)\, d\mu_t(x) \right| \longrightarrow 0 \quad \text{quand} \quad N \longrightarrow +\infty,$$

c'est-à-dire que

$$\mathbb{E}\left| \frac{1}{N} \sum_{i=1}^N \varphi(X_t^{i,N}) - \int_{\mathbb{R}^d} \varphi(x)\, d\mu_t(x) \right| \longrightarrow 0 \quad \text{quand} \quad N \longrightarrow +\infty$$

pour toute fonction test φ sur $\mathbb{R}^d$. Suite à ces résultats de type loi des grands nombres, les déviations des observables $\displaystyle\frac{1}{N} \sum_{i=1}^N \varphi(X_t^{i,N})$ autour de $\displaystyle\int_{\mathbb{R}^d} \varphi(x)\, d\mu_t(x)$ ont été étudiées dans les cadres asymptotiques d'un théorème central limite ou de grandes déviations (cf. [48], [84] ou [109] par exemple).

Il peut être intéressant, dans une perspective numérique, d'obtenir des estimations quantitatives de la convergence de $\hat{\mu}_t^N$ vers μ_t assurant que la méthode a de grandes chances de donner un bon résultat. Par exemple, si φ est une fonction test, peut-on mesurer l'écart entre l'observable $\displaystyle\frac{1}{N} \sum_{i=1}^N \varphi(X_t^{i,N})$ du système et la valeur $\displaystyle\int_{\mathbb{R}^d} \varphi(x)\, d\mu_t$ donnée par la solution μ_t de l'équation (33) en estimant par exemple la quantité

$$\mathbb{P}\left[\left| \frac{1}{N} \sum_{i=1}^N \varphi(X_t^{i,N}) - \int_{\mathbb{R}^d} \varphi(x)\, d\mu_t(x) \right| > \varepsilon \right]$$

en fonction de la taille N du système ? Notons en effet que l'utilisation de tels systèmes de particules stochastiques s'est révélée utile dans des méthodes numériques d'approximation d'équations de la forme de (33) (cf. [107] par exemple) : il s'agit alors de pouvoir évaluer l'erreur commise par la méthode numérique en fonction des différents paramètres tels que le nombre de particules ou le pas de temps de discrétisation des équations (35).

De tels résultats ont été obtenus récemment par F. Malrieu [74] en liaison avec les inégalités de concentration (ou de déviation) étudiées dans la partie II de cette introduction.

Se fondant sur un résultat de D. Bakry [7], il montre tout d'abord que la loi jointe $\mu_t^{(N)}$ du N-uplet $(X_t^{1,N},\dots,X_t^{N,N})$ vérifie une inégalité de Sobolev logarithmique de constante λ indépendante de t et N si μ_0 vérifie une telle inégalité et si les potentiels V et W vérifient certaines hypothèses de convexité. En particulier l'inégalité

$$\sup_{[\varphi]_{lip}\leq 1} \mathbb{P}\Big[\Big|\frac{1}{N}\sum_{i=1}^N \varphi(X_t^{i,N}) - \mathbb{E}\Big(\frac{1}{N}\sum_{i=1}^N \varphi(X_t^{i,N})\Big)\Big| > \varepsilon\Big] \leq 2\,e^{-\frac{\lambda}{2}N\varepsilon^2}$$

est vérifiée pour tout $\varepsilon > 0$. Il reprend ensuite l'argument de couplage présentée dans [105] par exemple, qui consiste en l'introduction d'une famille de processus indépendants $(Y_t^{i,N})_{t\geq 0}$, pour $1\leq i\leq N$, tels que $Y_t^{i,N}$ soit proche de $X_t^{i,N}$ et de loi μ_t en chaque instant t. Cette loi μ_t étant celle à l'instant t du processus $(Y_t)_{t\geq 0}$ solution de

$$dY_t = \sqrt{2}\,dB_t - \nabla V(Y_t) - \nabla W * \mu_t(Y_t)$$

pour une donnée initiale Y_0 de loi μ_0, les $(Y_t^{i,N})_{t\geq 0}$ sont définis comme les solutions des équations

$$dY_t^{i,N} = \sqrt{2}\,dB_t^{i,N} - \nabla V(Y_t^{i,N}) - \nabla W * \mu_t(Y_t^{i,N}), \qquad 1\leq i\leq N$$

pour les données initiales indépendantes $Y_0^{i,N} = X_0^{i,N}$ et où $(B_t^{i,N})_{t\geq 0}$ est le mouvement brownien dirigeant l'évolution de $X_t^{i,N}$ pour chaque i.

Il montre alors que les $X_t^{i,N}$ et les $Y_t^{i,N}$ restent proches en tout instant t au sens où

$$\mathbb{E}\,\big|X_t^{i,N} - Y_t^{i,N}\big|^2 \leq \frac{C^2}{N}$$

puis

$$\sup_{[\varphi]_{lip}\leq 1} \Big|\mathbb{E}\,\Big(\frac{1}{N}\sum_{i=1}^N \varphi(X_t^{i,N})\Big) - \mathbb{E}\,\Big(\frac{1}{N}\sum_{i=1}^N \varphi(Y_t^{i,N})\Big)\Big| \leq \frac{C}{\sqrt{N}}$$

pour une constante C indépendante du temps, puis en déduit le résultat suivant par inégalité triangulaire et hypothèses sur les $Y_t^{i,N}$:

Théorème 16 (cf. [74]). *Supposons que V soit de classe $\mathcal{C}^2$ sur $\mathbb{R}^d$ et uniformément convexe et que W soit de classe $\mathcal{C}^2$, convexe, paire et croisse de façon polynomiale à l'infini. Soient d'une part μ la solution de* (33) *de donnée initiale μ_0 satisfaisant une inégalité de Sobolev logarithmique sur $\mathbb{R}^d$ et d'autre part $(X_t^{i,N})_{t\geq 0}$ pour $1\leq i\leq N$ les solutions de* (35) *pour des données initiales $X_0^{i,N}$ indépendantes et de loi μ_0.*

Alors il existe deux constantes positives C et λ telles que

$$\sup_{t\geq 0}\ \sup_{[\varphi]_{lip}\leq 1} \mathbb{P}\Big[\Big|\frac{1}{N}\sum_{i=1}^N \varphi(X_t^{i,N}) - \int_{\mathbb{R}^d}\varphi(x)\,d\mu_t(x)\Big| > \frac{C}{\sqrt{N}} + \varepsilon\Big] \leq 2\,e^{-\frac{\lambda}{2}N\,\varepsilon^2}$$

pour tous $\varepsilon > 0$ et $N\geq 1$.

En particulier il existe deux constantes N_0 et λ telles que

$$\sup_{t\geq 0} \sup_{[\varphi]_{lip}\leq 1} \mathbb{P}\Big[\Big|\frac{1}{N}\sum_{i=1}^{N}\varphi(X_t^{i,N}) - \int_{\mathbb{R}^d}\varphi(x)\,d\mu_t(x)\Big| > \varepsilon\Big] \leq 2\,e^{-\frac{\lambda}{2}N\,\varepsilon^2} \tag{37}$$

pour tous $\varepsilon > 0$ et $N \geq N_0\,\varepsilon^{-2}$.

Un aspect de ce résultat est l'indépendance des constantes en t, sur laquelle nous reviendrons dans le paragraphe III.6. Nous nous proposons maintenant de prolonger cette étude en obtenant des estimations plus fortes sur la convergence de la mesure empirique.

III.4. Nouveaux résultats de déviation de la mesure empirique (cf. chapitre 6)

Dans le chapitre 6, sous une condition plus contraignante sur le nombre N de particules, nous renforçons les inégalités de déviation (37) sous la forme

$$\mathbb{P}\Big[\sup_{0\leq t\leq T} \sup_{[\varphi]_{lip}\leq 1} \Big|\frac{1}{N}\sum_{i=1}^{N}\varphi(X_t^{i,N}) - \int_{\mathbb{R}^d}\varphi(x)\,d\mu_t(x)\Big| > \varepsilon\Big] \leq C\,e^{-KN\varepsilon^2},$$

c'est-à-dire

$$\mathbb{P}\Big[\sup_{0\leq t\leq T} W_1(\hat{\mu}_t^N, \mu_t) > \varepsilon\Big] \leq C\,e^{-KN\varepsilon^2},$$

où W_1 est la distance de Wasserstein d'ordre 1 définie sur $\mathcal{P}_1(\mathbb{R}^d)$ à partir de la distance euclidienne sur $\mathbb{R}^d$ et pour certaines constantes positives C et K. Pour cela nous supposons que les potentiels V et W sont deux fois différentiables sur $\mathbb{R}^d$ et vérifient

$$D^2V(x) \geq \beta I, \quad \gamma I \leq D^2W(x) \leq \gamma' I, \qquad x \in \mathbb{R}^d \tag{38}$$

pour des constantes réelles β, γ, γ', et

$$|\nabla V(x)| = O(e^{a|x|^2}) \quad \text{quand} \quad |x| \to +\infty \tag{39}$$

pour tout $a > 0$. Sous ces hypothèses l'existence et l'unicité trajectorielles et en loi de solutions des équations stochastiques (34) et (35) peuvent être montrées en adaptant à notre cas la méthode développée par A.-S. Sznitman [105] dans le cas où ∇V et ∇W sont bornés et lipschitziens, puis étendue par S. Méléard au cas où ∇V et ∇W sont seulement lipschitziens ; le problème de Cauchy pour l'équation (33) est quant à lui traité dans [36].

Nous montrons alors le

Théorème 17 (cf. Theorem 6.7). *Supposons que V et W vérifient les hypothèses* (38) *et* (39). *Soient d'une part μ la solution de* (33) *de donnée initiale μ_0 admettant un moment exponentiel carré fini sur $\mathbb{R}^d$ et d'autre part $(X_t^{i,N})_{t\geq 0}$ pour $1 \leq i \leq N$ les solutions de* (35) *pour des données initiales $X_0^{i,N}$ indépendantes et de loi μ_0.*

Alors, notant $\hat{\mu}_t^N$ la mesure empirique des $X_t^{i,N}$, pour tout $T \geq 0$ il existe une constante K et pour tout $d' > d$ il existe deux constantes positives C et N_0 telles que

$$\mathbb{P}\Big[\sup_{0\leq t\leq T} W_1(\hat{\mu}_t^N, \mu_t) > \varepsilon\Big] \leq C\,(1 + T\varepsilon^{-2})\,e^{-K\,N\,\varepsilon^2} \tag{40}$$

pour tous $\varepsilon > 0$ et $N \geq N_0\,\max(\varepsilon^{-(d'+2)}, 1)$.

Nous montrons ainsi que non seulement la mesure empirique est proche de la mesure limite à chaque instant, de manière uniforme sur les observables, mais aussi que la probabilité d'observer une déviation significative sur l'intervalle de temps $[0, T]$ est petite. De plus, par rapport au théorème 16, nous étendons la classe de données initiales μ_0 admissibles à des mesures vérifiant uniquement une condition d'intégrabilité, cette condition étant impliquée par une inégalité de Sobolev logarithmique d'après la partie II.

En contre-partie nous nous limitons à des forces d'interaction ∇W lipschitziennes et imposons une condition (néanmoins explicite) sur la taille N du système de particules à considérer. Celle-ci apparaît dans la démonstration dont nous présentons maintenant l'idée.

Comme dans le paragraphe III.3 nous introduisons une famille de processus indépendants $(Y_t^{i,N})_{t\geq 0}$ pour $1 \leq i \leq N$, solutions de

$$dY_t^{i,N} = \sqrt{2}\, dB_t^{i,N} - \nabla V(Y_t^{i,N}) - \nabla W * \mu_t(Y_t^{i,N}), \qquad 1 \leq i \leq N$$

pour les données initiales $Y_0^{i,N} = X_0^{i,N}$ et où $(B_t^{i,N})_{t\geq 0}$ est le mouvement brownien dirigeant l'évolution de $X_t^{i,N}$ pour chaque i. Nous remplaçons l'estimation *en moyenne*

$$\mathbb{E}\big|X_t^{i,N} - Y_t^{i,N}\big|^2 \leq \frac{C^2}{N}$$

du paragraphe III.3, où dans notre cas C pourrait dépendre de t, par l'estimation *presque sûre*

$$\sup_{0\leq t\leq T} W_1(\hat{\mu}_t^N, \mu_t) \leq c \sup_{0\leq t\leq T} W_1(\hat{\nu}_t^N, \mu_t)$$

où $\hat{\nu}_t^N = \dfrac{1}{N}\displaystyle\sum_{i=1}^N \delta_{Y_t^{i,N}}$ est la mesure empirique des $Y_t^{i,N}$ et c une constante dépendant de T. En particulier

$$\mathbb{P}\Big[\sup_{0\leq t\leq T} W_1(\hat{\mu}_t^N, \mu_t) > \varepsilon\Big] \leq \mathbb{P}\Big[\sup_{0\leq t\leq T} W_1(\hat{\nu}_t^N, \mu_t) > \tilde{\varepsilon}\Big] \tag{41}$$

où $\tilde{\varepsilon} = \varepsilon/c$.

Dans un premier temps nous estimons la quantité $\mathbb{P}\big[W_1(\hat{\nu}_t^N, \mu_t) > \tilde{\varepsilon}\big]$. Le temps t étant fixé, il s'agit de mesurer la déviation de la mesure empirique de N variables indépendantes autour de leur loi commune. Pour cela, après avoir vérifié que μ_t admet un moment exponentiel carré par propagation de ce moment par l'équation, et donc satisfait une inégalité de transport T_1 d'après le théorème 9, nous déduisons du théorème 10 l'existence de constantes K et N_0, dépendant de t, telles que

$$\mathbb{P}\big[W_1(\hat{\nu}_t^N, \mu_t) > \tilde{\varepsilon}\big] \;\leq\; e^{-K\,N\,\tilde{\varepsilon}^2}$$

pour tous $\tilde{\varepsilon} > 0$ et $N \geq N_0 \max(\tilde{\varepsilon}^{-(d'+2)}, 1)$, puis ensuite une borne pour

$$\mathbb{P}\big[\sup_{0\leq t\leq T} W_1(\hat{\nu}_t^N, \mu_t) > \tilde{\varepsilon}\big]$$

grâce à des propriétés de continuité en temps de μ_t et des $Y_t^{i,N}$. Nous concluons alors la démonstration de l'inégalité (40) grâce à (41).

En termes de propagation du chaos, si par exemple

$$\hat{\mu}_t^{N,2} = \frac{1}{N(N-1)} \sum_{i \neq j} \delta_{(X_t^{i,N}, X_t^{j,N})}$$

est la mesure empirique des *paires* de particules, l'indépendance asymptotique de deux particules quelconques du système est donnée par le résultat suivant, qui se démontre de manière analogue au théorème 17 :

Théorème 18 (cf. Theorem 6.8). *Sous les hypothèses du théorème 17, pour tous $T \geq 0$ et $d' > d$ il existe deux constantes positives K et N_0 telles que*

$$\mathbb{P}\Big[\sup_{0 \leq t \leq T} W_1(\hat{\mu}_t^{N,2}, \mu_t \otimes \mu_t) > \varepsilon\Big] \leq e^{-K N \varepsilon^2}$$

pour tous $\varepsilon > 0$ et $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$.

Ici W_1 désigne la distance de Wasserstein d'ordre 1 définie sur $\mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$ à partir de la distance euclidienne sur $\mathbb{R}^d \times \mathbb{R}^d$.

Une application possible d'une variante du théorème 17 sera donnée dans le paragraphe III.6.

III.5. Extension des résultats au niveau des trajectoires (cf. chapitre 7)

Les théorèmes 16 et 17 donnent des estimations de déviation de la mesure empirique $\hat{\mu}_t^N$ des particules au temps t autour de la valeur μ_t au temps t de la solution de l'équation (33), c'est-à-dire autour de la loi au temps t du processus $(Y_t)_{t \geq 0}$ défini par

$$dY_t = \sqrt{2}\, dB_t - \nabla V(Y_t) - \nabla W * \mu_t(Y_t) \tag{42}$$

pour une donnée initiale Y_0 de loi μ_0.

Dans le chapitre 7 nous établissons des résultats analogues au niveau des *trajectoires* des particules. Pour cela, étant donné $T \geq 0$, nous considérons la mesure empirique des trajectoires $(X_t^{i,N})_{0 \leq t \leq T}$ sur l'intervalle de temps $[0, T]$, définie par

$$\hat{\mu}_{[0,T]}^N = \frac{1}{N} \sum_{i=1}^N \delta_{(X_t^{i,N})_{0 \leq t \leq T}}.$$

Chaque trajectoire étant un élément de l'espace $\mathcal{C}$ des fonctions continues de $[0, T]$ dans $\mathbb{R}^d$, il s'agit d'une mesure aléatoire sur $\mathcal{C}$ et non plus sur $\mathbb{R}^d$ comme précédemment.

Nous étendons les estimations de déviation (40) à ce nouveau cadre (qui est le cadre utilisé par de nombreux auteurs) :

Théorème 19 (cf. Theorem 7.2). *Supposons que ∇V et ∇W soient lipschitziens sur $\mathbb{R}^d$ et soit μ_0 une mesure de probabilité sur $\mathbb{R}^d$ admettant un moment exponentiel carré fini. Etant donné $T \geq 0$, soit d'une part $\mu_{[0,T]}$ la loi du processus solution de (42) sur $[0,T]$ pour une donnée initiale de loi μ_0 et soit d'autre part $\hat{\mu}^N_{[0,T]}$ la mesure empirique des solutions $(X^{i,N}_t)_{0\leq t\leq T}$ de (35) sur $[0,T]$ pour des données initiales $X^{i,N}_0$ indépendantes et de loi μ_0.*

Alors, pour tout $\alpha \in (0,1/2)$, il existe deux constantes positives K et N_0 telles que

$$\mathbb{P}\Big[\sup_{0\leq t\leq T} W_1(\hat{\mu}^N_t, \mu_t) > \varepsilon\Big] \leq e^{-K\,N\,\varepsilon^2} \tag{43}$$

pour tous $\varepsilon > 0$ et $N \geq N_0\,\varepsilon^{-2}\,\exp\big(N_0\,\varepsilon^{-1/\alpha}\big)$.

Ici W_1 désigne la distance de Wasserstein d'ordre 1 définie sur $\mathcal{P}_1(\mathcal{C})$ à partir de la norme uniforme sur $\mathcal{C}$.

Par projection au temps t nous retrouvons les inégalités de déviation (40) pour les marginales en temps (sous une condition plus forte sur la taille du système de particules). Mais surtout nous obtenons des estimations de concentration au niveau des trajectoires dans leur ensemble. Nous espérons qu'elles se révéleront utiles pour de futures applications faisant intervenir la probabilité que le système soit dans un état A au temps s puis dans un état B au temps t, et qui, nous semble-t-il, ne peuvent être traitées par des estimations portant uniquement sur les marginales en temps.

Le théorème 19 se démontre par une adaptation de la preuve du théorème 17, en utilisant en particulier le théorème 11 à la place du théorème 10. Une comparaison des deux démonstrations laisse apparaître qu'elle est plus simple dans le sens où elle comporte moins d'étapes, mais qu'en contre-partie celles-ci sont plus délicates : par exemple, pour vérifier que la loi $\mu_{[0,T]}$ satisfait les hypothèses du théorème 11, nous faisons appel à des propriétés fortes de régularité et d'intégrabilité des solutions d'équations différentielles stochastiques.

Dans le prochain paragraphe nous verrons que le théorème 17 peut être précisé dans certains cas où une solution d'équilibre de l'équation (33) a été mise en évidence.

III.6. Approximation particulaire des solutions stationnaires (cf. chapitre 6)

Les solutions de l'équation (33) ont récemment été étudiées pour leur comportement asymptotique quand t tend vers l'infini.

Tout d'abord, D. Benedetto, E. Caglioti, J. A. Carrillo et M. Pulvirenti [11] ont établi, dans le cas où $V \equiv 0$ et $W(x) = |x|^3/3$ en dimension 1, l'existence d'un profil stationnaire μ_∞ et la convergence des solutions μ_t vers ce profil quand le temps t tend vers l'infini. Puis l'équation (33) a été interprétée comme un flot gradient sur la mesure solution dans la structure différentielle développée sur l'espace des mesures de probabilité (de second moment fini) par F. Otto et mentionnée dans la partie I de cette introduction. L'énergie définissant ce flot gradient est l'énergie libre définie par

$$F(\nu) = \int_{\mathbb{R}^d} f(x)\,\ln f(x)\,dx + \int_{\mathbb{R}^d} V(x)\,f(x)\,dx + \frac{1}{2}\iint_{\mathbb{R}^{2d}} W(x-y)\,f(x)\,f(y)\,dx\,dy$$

si ν est absolument continue par rapport à la mesure de Lebesgue, de densité f. Grâce à cette interprétation, et s'appuyant sur des techniques de dissipation d'entropie liées aux travaux de D. Bakry et M. Emery sur les inégalités de Sobolev logarithmiques et fondées sur une comparaison des dérivées successives de $F(\mu_t)$ par rapport à t, J. A. Carrillo, R. McCann et C. Villani [36, 37] ont montré des résultats de convergence exponentielle en temps des solutions μ_t de l'équation vers un profil limite μ_∞ minimisant l'énergie libre F. Par exemple, pour des potentiels V et W vérifiant (38) avec $\beta > 0, \beta + 2\,\gamma > 0$, ils obtiennent un résultat de la forme

$$W_2(\mu_t, \mu_\infty) \leq C\, e^{-(\beta+2\,\gamma)t}, \qquad t > 0$$

pour une constante C dépendant de μ_0.

Certains de ces résultats de convergence ont été obtenus de manière indépendante par F. Malrieu [74] grâce à l'approximation particulaire (35) de (33).

Dans ce cadre nous pouvons espérer que la mesure empirique $\hat{\mu}_t^N$ du système au temps t soit une bonne approximation de la solution stationnaire μ_∞ quand N et t tendent vers l'infini, uniformément en temps. En effet dans le chapitre 6 nous précisons le théorème 17 sous la forme suivante :

Théorème 20 (cf. Theorem 6.9). *Avec les notations et sous les hypothèses du théorème 17, supposons que $\beta > 0$, et $\beta + 2\,\gamma > 0$. Alors*

(i) il existe une constante positive K telle que, pour tout $d' > d$, il existe des constantes C et N_0 telles que

$$\sup_{t\geq 0} \mathbb{P}\left[W_1(\hat{\mu}_t^N, \mu_t) > \varepsilon\right] \leq C\,(1+\varepsilon^{-2})\, e^{-K\,N\,\varepsilon^2}$$

pour tous $\varepsilon > 0$ et $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$;

(ii) il existe de plus des constantes T_0 et ε_0 telles que

$$\sup_{t\geq T_0 \ln(\varepsilon_0/\varepsilon)} \mathbb{P}\left[W_1(\hat{\mu}_t^N, \mu_\infty) > \varepsilon\right] \leq C\,(1+\varepsilon^{-2})\, e^{-\frac{K}{4}\,N\,\varepsilon^2}$$

pour tous $\varepsilon > 0$ et $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$.

Les hypothèses de convexité imposées à V et W permettent l'obtention de telles estimations uniformes en temps, comme nous l'avons noté dans le théorème 16.

Dans ce qui précède nous avons approché la solution au temps t de l'équation (33), ou son profil limite, par une mesure empirique $\hat{\mu}_t^N$ au sens de la distance de Wasserstein W_1 qui, sous une condition de moments, métrise seulement la convergence faible (étroite) des mesures de probabilité. Or sous nos hypothèses sur V et W, les mesures μ_∞ et μ_t pour $t > 0$ ont une densité régulière par rapport à la mesure de Lebesgue (cf. **Proposition 6.19**). Nous pouvons alors chercher à approcher ces densités de manière plus forte, en norme uniforme par exemple.

Pour cela nous régularisons la mesure empirique $d\hat{\mu}_t^N(x)$ en $\hat{f}_t^{N,\alpha}(x)\,dx$ avec

$$\hat{f}_t^{N,\alpha}(x) = \frac{1}{N}\sum_{i=1}^N \zeta_\alpha\big(x - X_t^i\big), \tag{44}$$

où $\zeta_\alpha = \alpha^{-d}\zeta(./\alpha)$ pour une fonction ζ positive et de classe $\mathcal{C}^\infty$ sur $\mathbb{R}^d$, à support compact et d'intégrale 1. Ainsi par exemple, dans le chapitre 6, nous montrons et utilisons des propriétés de régularité de la densité des μ_t pour établir le résultat suivant sur la densité f_∞ de la mesure d'équilibre μ_∞ :

Théorème 21 (cf. Theorem 6.11). *Avec les notations et sous les hypothèses du théorème 20, pour tout $\varepsilon > 0$ il existe un $\alpha = O(\varepsilon)$ tel que la mesure empirique régularisée définie par* (44) *vérifie*

$$\sup_{t \geq T_0 \ln(\varepsilon_0/\varepsilon)} \mathbb{P}\left[\|\hat{f}_t^{N,\alpha} - f_\infty\|_{L^\infty(\mathbb{R}^d)} > \varepsilon\right] \leq C\,(1 + \varepsilon^{-(2d+4)})\, e^{-\frac{K}{4}\,N\,\varepsilon^{2d+4}}$$

pour tout $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$.

Dans tous ces énoncés, les constantes peuvent être calculées explicitement en fonction des données.

Voici donc donnés quelques résultats d'approximation de la solution de l'équation (33) par une méthode particulaire.

Il pourrait être intéressant d'affaiblir les hypothèses faites sur les potentiels V et W, afin de pouvoir inclure en particulier le cas intéressant de l'équation des milieux granulaires, qui a été une motivation de cette étude, et pour lequel $W(x) = |x|^3/3$.

Diverses extensions de ce travail sont envisageables : par exemple, à un niveau théorique, considérer le cas où les données initiales ne sont plus indépendantes, mais constituent un processus de Markov par exemple, ce qui nous ramène en particulier à des problèmes d'inégalités de déviation pour des variables dépendantes, en partie abordés dans la partie II de cette introduction ; ou également, dans une optique numérique, étudier les erreurs supplémentaires commises dans la discrétisation en temps des équations décrivant l'évolution des particules en interaction.

Première partie

Propriétés de type contraction de certaines équations aux dérivées partielles

Chapitre 1

Propriétés de stabilité exponentielle et limite de champ moyen pour l'équation de Vlasov

Notant que le problème de la limite de champ moyen pour l'équation de Vlasov peut se ramener à un problème de stabilité de solutions de l'équation, nous nous intéressons à des propriétés de type contraction, ou plutôt d'expansion contrôlée, pour certaines distances mesurant l'écart entre solutions. En particulier nous montrons que l'utilisation des distances de Wasserstein d'ordre 1 *et surtout* 2*, pour laquelle nous disposons d'un cadre agréable développé récemment, permettent de mieux contrôler la croissance de la distance entre solutions.*

Introduction

Let us consider the Vlasov equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f - \nabla_x V * \rho[f_t] \cdot \nabla_v f = 0, \qquad t > 0, \quad x, v \in \mathbb{R}^d. \tag{1.1}$$

where the unknown $f_t : (x, v) \mapsto f(t, x, v)$ represents the density of presence of a system at time t in the phase space $\mathbb{R}^d \times \mathbb{R}^d$; in the equation the density in the space of positions $\rho[f_t]$ and the force field $\nabla_x V * \rho[f_t]$ are given by

$$\rho[f_t](x) = \int_{\mathbb{R}^d} f_t(x, v)\, dv, \qquad x \in \mathbb{R}^d$$

and

$$\nabla_x V * \rho[f_t](x) = \int_{\mathbb{R}^d} \nabla_x V(x - y)\, \rho[f_t](y)\, dy, \qquad x \in \mathbb{R}^d.$$

We also let $a \cdot b$ denote the scalar product of two vectors a and b in $\mathbb{R}^d$, and ∇_x or ∇_v denote the gradient operator in $\mathbb{R}^d$.

We shall assume that the potential V is symmetric in the sense that $V(-x) = V(x)$ for any $x \in \mathbb{R}^d$, and differentiable with Lipschitz gradient $\nabla_x V$. In the case when $\nabla_x V$ is moreover bounded, existence and uniqueness of measure-valued solutions to (1.1) have been proven in [25], [85] and [104] : more precisely, given f_0 in the space $\mathcal{P}(\mathbb{R}^{2d})$ of probability measures on $\mathbb{R}^{2d}$, there exists a unique map in $\mathcal{C}([0,+\infty[, \mathcal{P}(\mathbb{R}^{2d}))$, where $\mathcal{P}(\mathbb{R}^{2d})$ is equipped with the usual weak (narrow) topology, such that (1.1) hold in the sense of distributions with the initial datum f_0.

1.1 A first stability result and the particle approximation

The proof is based on some estimate involving the dual-bounded Lipschitz distance d_{BL} defined on $\mathcal{P}(\mathbb{R}^{2d})$ by

$$d_{BL}(\mu, \nu) = \sup_{\|\varphi\|_{lip} \leq 1} \left\{ \int_{\mathbb{R}^{2d}} \varphi(z)\, d\mu(z) - \int_{\mathbb{R}^{2d}} \varphi(z)\, d\nu(z) \right\}$$

where $\| \, . \, \|_{lip}$ is the Lipschitz norm defined as

$$\|\varphi\|_{lip} = \max \left[\sup_z |\varphi(z)|, \sup_{w \neq z} \frac{|\varphi(w) - \varphi(z)|}{\|w - z\|_{\ell^1}} \right].$$

and $\| \, . \, \|_{\ell^1}$ is the norm on $\mathbb{R}^{2d}$ defined by $\|(z_1, z_2)\|_{\ell^1} = |z_1| + |z_2|$ where $|z_i|$ is the Euclidean norm of z_i in $\mathbb{R}^d$. The dual-bounded Lipschitz distance metrizes the weak (narrow) topology on $\mathcal{P}(\mathbb{R}^{2d})$, as pointed out in the appendix or in [47].

The uniqueness result goes in [104] with a really strong stability result relatively to the distance d_{BL}, which can be stated as follows :

Theorem 1.1 (cf. [104]). *Assume that ∇V is bounded by some constant B and Lipschitz with Lipschitz seminorm L. If f and g are solutions to* (1.1) *with initial data f_0 and g_0 in $\mathcal{P}(\mathbb{R}^{2d})$, then*

$$d_{BL}(f_t, g_t) \leq e^{c\,t}\, d_{BL}(f_0, g_0) \tag{1.2}$$

for any $t \geq 0$, where $c = (2 \max(B, 1) + 1) \max(L, 1)$.

Let us now explain how one can solve the mean field limit for this Vlasov equation from this estimate.

On one hand let f be a given solution to the Vlasov equation with initial datum f_0.

On the other hand let us consider N particles in the phase space $\mathbb{R}^{2d}$, with initial positions (X_0^i, V_0^i) evolving into (X_t^i, V_t^i), for $1 \leq i \leq N$ according to the Newton equations

$$\left\{ \begin{array}{rcl} \dfrac{dX_t^i}{dt} & = & V_t^i \\ \dfrac{dV_t^i}{dt} & = & -\dfrac{1}{N} \displaystyle\sum_{j=1}^{N} \nabla_x V(X_t^j - X_t^i) \end{array} \right. \qquad 1 \leq i \leq N.$$

Then one can note that the empirical measure

$$\hat{\mu}_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{(X_t^i, V_t^i)}$$

of the system is a solution to the Vlasov equation (see [25]).

Since the same holds for f, Theorem 1.1 ensures that

$$d_{BL}(f_t, \hat{\mu}_t^N) \leq e^{ct}\, d_{BL}(f_0, \hat{\mu}_0^N).$$

If in particular the initial positions (X_0^i, V_0^i) are such that $\hat{\mu}_0^N$ be close to f_0 in the weak (narrow) sense of probability measures, so that $d_{BL}(f_0, \hat{\mu}_0^N)$ is small, then $\hat{\mu}_t^N$ will remain close to f_t at later times : more precisely, if $d_{BL}(f_0, \hat{\mu}_0^N) = \varepsilon(N)$, then $d_{BL}(f_t, \hat{\mu}_t^N)$ will remain of order $\varepsilon(N)^\beta$ with $\beta < 1$, uniformly on a time interval of order $|\ln \varepsilon(N)|$ (see also [32] for instance for the approximation of stable stationary solutions by means of particles initially drawn in a random way).

1.2 Improving the stability results by using Wasserstein distances

In this section we aim at getting stability results analogous to (1.2) for stronger distances and with a better control on the growth on the constants.

First of all one can adapt the argument developed in [104] for the distance d_{BL} to the Wasserstein distance W_1 of order 1 defined from the norm $\| \,.\, \|_{\ell^1}$ on $\mathbb{R}^{2d}$, on the space $\mathcal{P}_1(\mathbb{R}^{2d})$ of probability measures on $\mathbb{R}^{2d}$ with finite first moment. Note that $W_1 \geq d_{BL}$ since, by the Kantorovich-Rubinstein dual formulation,

$$W_1(\mu, \nu) = \sup_{[\varphi]_{lip} \leq 1} \left\{ \int_{\mathbb{R}^{2d}} \varphi(z)\, d\mu(z) - \int_{\mathbb{R}^{2d}} \varphi(z)\, d\nu(z) \right\}$$

where $[\,.\,]_{lip}$ is the Lipschitz seminorm defined from the $\| \,.\, \|_{\ell^1}$ norm on $\mathbb{R}^{2d}$.

Then, as soon as the potential V is symmetric and has Lipschitz gradient, one can prove existence and uniqueness of solutions to (1.1) in the space $\mathcal{P}_1(\mathbb{R}^{2d})$. In particular the propagation of the first moment can be checked by the following a priori estimate.

Let f be a solution to the Vlasov equation (1.1), and let $<x>$ stand for $(1 + |x|^2)^{1/2}$ for x in $\mathbb{R}^d$. Then

$$\begin{aligned} \frac{d}{dt}\int_{\mathbb{R}^{2d}} (<x> + <v>)\, df_t(x,v) &= \int_{\mathbb{R}^{2d}} \Big(\frac{x}{<x>} \cdot v - \frac{v}{<v>} \cdot \nabla_x V * \rho[f_t](x) \Big)\, df_t(x,v) \\ &\leq \int_{\mathbb{R}^{2d}} (|v| + |\nabla_x V * \rho[f_t](x)|)\, df_t(x,v). \end{aligned}$$

But $\nabla_x V$ is L-Lipschitz and equal to 0 at 0, so $|\nabla_x V(x-y)| \leq L|x-y| \leq L(|x| + |y|)$ for any x and y in $\mathbb{R}^d$, and

$$\int_{\mathbb{R}^{2d}} |\nabla_x V * \rho[f_t](x)|\, df_t(x,v) \leq \iint_{\mathbb{R}^{4d}} |\nabla_x V(x-y)|\, df_t(x,v)\, df_t(y,w) \leq 2L \int_{\mathbb{R}^{2d}} |x|\, df_t(x,v).$$

Hence

$$\frac{d}{dt}\int_{\mathbb{R}^{2d}} (<x> + <v>)\, df_t(x,v) \leq \max(1, 2L) \int_{\mathbb{R}^{2d}} (<x> + <v>)\, df_t(x,v)$$

and then

$$\int_{\mathbb{R}^{2d}} (<x> + <v>)\, df_t(x,v) \leq e^{\max(1,2L)\, t} \int_{\mathbb{R}^{2d}} (<x> + <v>)\, df_0(x,v).$$

Now $\displaystyle\int_{\mathbb{R}^{2d}} (<x> + <v>)\, df_0(x,v)$ is finite if f_0 belongs to $\mathcal{P}_1(\mathbb{R}^{2d})$, and consequently so is $\displaystyle\int_{\mathbb{R}^{2d}} (<x> + <v>)\, df_t(x,v)$ for any $t \geq 0$. In other words f_t also belongs to $\mathcal{P}_1(\mathbb{R}^{2d})$.

Furthermore the stability result given in Theorem 1.1 turns into

Theorem 1.2. *Assume that $\nabla_x V$ is Lipschitz with Lipschitz seminorm L. If f and g are solutions to* (1.1) *with initial data f_0 and g_0 in $\mathcal{P}_1(\mathbb{R}^{2d})$, then*

$$W_1(f_t, g_t) \leq e^{ct}\, W_1(f_0, g_0)$$

for any $t \geq 0$, where $c = L + \max(L, 1)$.

We shall not give the proof of this result since it is just an adaptation of the proof of Theorem 1.1 given in [104], and is detailled in [111, Problem 14] with $c = 2\max(L,1)$ as a constant.

Instead we want to obtain such a result for the Wasserstein distance W_2 of order 2, defined from the usual Euclidean norm on $\mathbb{R}^{2d}$, on the space $\mathcal{P}_2(\mathbb{R}^{2d})$ of probability measures μ on $\mathbb{R}^{2d}$ with finite moment of order 2.

Let us note that, as a general fact, amongst all Wasserstein distances W_p, the distances W_1 and W_2 have been particularly used to study issues arising in probability theory, partial differential equations, etc. Indeed W_1 satisfies the convenient Kantorovich-Rubinstein dual formulation : in particular bounding by above the distance $W_1(\mu, \nu)$ for some measures μ and ν on $\mathbb{R}^d$ amounts to bounding the integral $\displaystyle\int_{\mathbb{R}^d} \varphi(x)\, d(\mu - \nu)(x)$ uniformly over the 1-Lipschitz functions. On the other hand the space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ has been equipped with a differential structure, first in [87], motivated by issues in PDE problems, then in more detail in [2], [3] or [37] for instance. This structure has revealed really adapted to the study of numerous PDE's such as the porous medium equation or some granular media equation, which in this structure can be seen as gradient flows on the measure solution of the equation (see for instance [36], [37], [87] and [111, Chapter 8]). In this formalism the Vlasov equation can be seen as a Hamiltonian equation on the measure; however we shall not use this fact to obtain our stability estimates.

In this work we just want to see, through the example of the Vlasov equation, that using the W_2 distance can enable to refine a result obtained, in a maybe simpler way, with the W_1 distance : more precisely we are going to improve the $\exp[(L + \max(L, 1))t]$ term obtained in Theorem 1.2 into a $\exp[(1 + L)t/2]$ term, proving in a formal way the following

Theorem 1.3. *Assume that V is twice differentiable, with Hessian matrix such that $-L\,I \leq D_x^2 V(x) \leq L\,I$ for some constant L and any x in $\mathbb{R}^d$. If f and g are solutions to* (1.1) *with initial data f_0 and g_0 in $\mathcal{P}_2(\mathbb{R}^{2d})$, then*

$$W_2(f_t, g_t) \leq e^{\left(\frac{1}{2}+L\right)\,t}\, W_2(f_0, g_0)$$

for any $t \geq 0$.

If moreover f_0 and g_0 have same center of mass and mean velocity, then

$$W_2(f_t, g_t) \leq e^{\frac{1+L}{2}\,t}\, W_2(f_0, g_0)$$

for any $t \geq 0$.

By center of mass and mean velocity of a measure μ on $\mathbb{R}^{2d}$ we mean the quantities $\displaystyle\int_{\mathbb{R}^{2d}} x\, d\mu(x,v)$ and $\displaystyle\int_{\mathbb{R}^{2d}} v\, d\mu(x,v)$.

Proof. Unless specified all integrals are meant in $\mathbb{R}^{2d}$ and the gradient $\nabla = (\nabla_x, \nabla_v)$ is meant with respect to the couple $z = (x,v)$ in $\mathbb{R}^{2d}$. The scalar product between two vectors a and b in $\mathbb{R}^d$ (resp. $\mathbb{R}^{2d}$) is denoted $a \cdot b$ (resp. $<a,b>$) ; the corresponding Euclidean norm is denoted $|\,.\,|$ (resp. $\|\,.\,\|$) and the divergence of a vector field U in $\mathbb{R}^{2d}$ is denoted $<\nabla, U>$.

We first note that any solution f to the Vlasov equation with initial datum f_0 in $\mathcal{P}_2(\mathbb{R}^{2d})$ is such that f_t also belongs to $\mathcal{P}_2(\mathbb{R}^{2d})$ for any $t \geq 0$. For this it is sufficient to check that the propagation of the moment of order 2 holds, namely that $\displaystyle\int (|x|^2 + |v|^2)\, df_t(x,v)$ is finite if so is $\displaystyle\int (|x|^2 + |v|^2)\, df_0(x,v)$, which can be done along the argument given above for the first moment.

Let us now write the Vlasov equation as

$$\frac{\partial f}{\partial t} + <\nabla, u[f_t]\, f> = 0, \qquad t > 0, \quad z \in \mathbb{R}^{2d} \tag{1.3}$$

where

$$u[h](z) = (v, -\nabla_x V * \rho[h](x))$$

for any $z = (x,v)$ in $\mathbb{R}^{2d}$ and any probability measure h on $\mathbb{R}^{2d}$.

Let $t \geq 0$ be given, and assume that f_t and g_t are absolutely continuous with respect to the Lebesgue measure. Then, as we have seen in the first part of the introduction, there exists some function φ_t on $\mathbb{R}^{2d}$ such that the map $\nabla\varphi_t$ send the measure f_t onto g_t in an optimal way (for W_2), in the sense that g_t is the image measure $\nabla\varphi_t \sharp f_t$ of f_t by $\nabla\varphi_t$, and

$$W_2^2(f_t, g_t) = \int \|\nabla\varphi_t(z) - z\|^2\, df_t(z).$$

Moreover, letting φ_t^* be the Legendre transform of the convex function φ_t, the map $\nabla\varphi_t^*$ sends the measure g_t onto f_t in an optimal way. Note that these facts are independent of the dynamics.

But, by (1.3), a result by L. Ambrosio, N. Gigli and G. Savaré, given in [2, Theorem 3. 12] (see also [3] or [111, Section 8.5]), ensures that the squared W_2 distance between f_t and g_t is derivable with respect to time, with

$$\frac{1}{2}\frac{d}{dt}W_2^2(f_t, g_t) = \int < z - \nabla\varphi_t^*(z), u[g_t](z) > \, dg_t(z) - \int < \nabla\varphi_t(z) - z, u[f_t](z) > \, df_t(z). \quad (1.4)$$

Now we aim at bounding the right hand side in terms of $W_2^2(f_t, g_t)$ so as to conclude the argument by Gronwall's lemma.

For this we now let t be fixed and, for notational convenience, let $\varphi = \varphi_t$, $\mu_0 = f_t$ and $\mu_1 = g_t$. Given $s \in [0, 1]$ we also let $T^s = (1 - s)\, Id + s\, \nabla\varphi$ and consider the displacement interpolation between μ_0 and μ_1 defined by $\mu_s = T^s \sharp \mu_0$. Then it can be checked that μ_s is a weak solution to the continuity equation

$$\frac{\partial \mu_s}{\partial s} + < \nabla, \mu_s\, v_s > = 0, \qquad s \in [0, 1], \quad z \in \mathbb{R}^{2d}$$

where the velocity v_s is given by

$$v_s(T^s(z)) = \frac{d}{ds}T^s(z) = \nabla\varphi(z) - z.$$

In this notation, the right hand side in (1.4) is equal to

$$\int < v_1(z), u[\mu_1](z) > \, d\mu_1(z) - \int < v_0(z), u[\mu_0](z) > \, d\mu_0(z)$$

$$= F(1) - F(0) = \int_0^1 F'(s)\, ds \quad (1.5)$$

where, for $s \in [0, 1]$,

$$F(s) = \int < v_s(z), u[\mu_s](z) > \, d\mu_s(z) = \int < \theta(z), u[\mu_s](T^s(z)) > \, d\mu_0(z)$$

with $\theta(z) = \nabla\varphi(z) - z$. Note that θ is the displacement made when transporting μ_0 onto μ_1. By derivation

$$F'(s) = \int < \theta(z), \frac{d}{ds}\big(u[\mu_s](T^s(z))\big) > \, d\mu_0(z). \quad (1.6)$$

We introduce again some notation. Given z in $\mathbb{R}^{2d}$ we decompose $\theta(z)$ as $(\theta_x(z), \theta_v(z))$ with both components in $\mathbb{R}^d$; in the same way we let $T^s(z) = (T_x^s(z), T_v^s(z))$ and have already let $z = (x, v)$ and $\nabla\varphi(z) = (\nabla_x\varphi(z), \nabla_v\varphi(z))$.

In this notation, given $s \in [0, 1]$, the term $u[\mu_s](T^s(z)) = u[\mu_s](z + s\, \theta(z))$ is the vector $\big(v + s\, \theta_v(z), -\nabla_x V * \rho[\mu_s](T_x^s(z))\big)$. But

$$\nabla_x V * \rho[\mu_s](T_x^s(z)) = \int \nabla_x V(T_x^s(z) - T_x^s(w))\, d\mu_0(w),$$

so
$$\frac{d}{ds}\big(\nabla_x V * \rho[\mu_s](T_x^s(z))\big) = \int D_x^2 V(T_x^s(z) - T_x^s(w))(\theta_x(z) - \theta_x(w))\, d\mu_0(w).$$
Hence
$$\frac{d}{ds}\, u[\mu_s](T^s(z)) = \Big(\theta_v(z), -\int D_x^2 V(T_x^s(z) - T_x^s(w))(\theta_x(z) - \theta_x(w))\, d\mu_0(w)\Big)$$
and by (1.6)
$$F'(s) = \int \theta_x(z)\cdot\theta_v(z)\, d\mu_0(z) - \iint_{\mathbb{R}^{4d}} \theta_v(z)\cdot D_x^2 V(T_x^s(z) - T_x^s(w))(\theta_x(z) - \theta_x(w))\, d\mu_0(w)\, d\mu_0(z). \tag{1.7}$$

By Young's inequality and optimality of φ between μ_0 and μ_1 the first term is bounded by
$$\frac{1}{2}\int |\theta_x(z)|^2 + |\theta_v(z)|^2\, d\mu_0(z) = \frac{1}{2}\int \|\nabla\varphi(z) - z\|^2\, d\mu_0(z) = \frac{1}{2} W_2^2(\mu_0, \mu_1). \tag{1.8}$$

Since V is a symmetric function, the second term in (1.7) is equal to
$$\frac{1}{2}\iint_{\mathbb{R}^{4d}} (\theta_v(z) - \theta_v(w)) \;\cdot\; D_x^2 V(T_x^s(z) - T_x^s(w))(\theta_x(z) - \theta_x(w))\, d\mu_0(w)\, d\mu_0(z)$$
and then, since moreover V has Hessian matrix bounded by $L\, I$, it is bounded by
$$\begin{aligned} &\frac{L}{2}\iint_{\mathbb{R}^{4d}} |\theta_v(z) - \theta_v(w)|\, |\theta_x(z) - \theta_x(w)|\, d\mu_0(w)\, d\mu_0(z) \\ &\le \frac{L}{2}\Big(\iint_{\mathbb{R}^{4d}} |\theta_v(z) - \theta_v(w)|^2\, d\mu_0(w)\, d\mu_0(z)\Big)^{1/2}\Big(\iint_{\mathbb{R}^{4d}} |\theta_x(z) - \theta_x(w)|^2\, d\mu_0(w)\, d\mu_0(z)\Big)^{1/2} \end{aligned} \tag{1.9}$$

by Cauchy-Schwarz inequality.

In the general case when f_0 and g_0 are not assumed to have same center of mass and mean velocity, (1.9) is bounded by
$$2\,L\Big(\int |\theta_x(z)|^2\, d\mu_0(z)\Big)^{1/2}\Big(\int |\theta_v(z)|^2\, d\mu_0(z)\Big)^{1/2} \le L\int |\theta_x(z)|^2 + |\theta_v(z)|^2\, d\mu_0(z) = L W_2^2(\mu_0, \mu_1)$$
by Young's inequality and the identities in (1.8).

Hence, by (1.7), (1.8) and (1.9) we have obtained the bound
$$F'(s) \le \Big(\frac{1}{2} + L\Big)\, W_2^2(\mu_0, \mu_1)$$
for any $s \in [0, 1]$ so that, in the original notation,
$$\frac{1}{2}\frac{d}{dt} W_2^2(f_t, g_t) \le \Big(\frac{1}{2} + L\Big)\, W_2^2(f_t, g_t)$$

by (1.4) and (1.5).

Then one can conclude the argument of the first statement in Theorem 1.3 since this holds for any t.

Let us now assume that the initial data f_0 and g_0 have same center of mass and mean velocity. Since V is a symmetric function, this property is conserved by the Vlasov equation. In particular the assumptions of Lemma 1.4 below hold with $\mu_0 = f_t$ and $\mu_1 = g_t$ and $\theta = \varphi_t - Id$. This implies that (1.9) is now bounded by

$$L\Big(\int |\theta_v(z)|^2\, d\mu_0(z)\Big)^{1/2}\Big(\int |\theta_x(z)|^2\, d\mu_0(z)\Big)^{1/2} \leq \frac{L}{2}\int |\theta_v(z)|^2 + |\theta_x(z)|^2\, d\mu_0(z) = \frac{L}{2}W_2^2(\mu_0, \mu_1)$$

by Young's inequality and optimality of $Id + \theta$ in the transport between μ_0 and μ_1 again.

Then one can conclude the proof by following the argument given in the general case. □

In this proof we have used the following general

Lemma 1.4. *Let μ_0 and μ_1 be in $\mathcal{P}_2(\mathbb{R}^{2d})$ with same center of mass and mean velocity, and let θ be a map from $\mathbb{R}^{2d}$ into itself such that $\mu_1 = (Id + \theta)\sharp\mu_0$. Then*

$$\iint_{\mathbb{R}^{4d}} |\theta_i(z) - \theta_i(w)|^2\, d\mu_0(z)\, d\mu_0(w) = 2\int_{\mathbb{R}^{2d}} |\theta_i(z)|^2\, d\mu_0(z)$$

where i stands for x or v, and $\theta(z) = (\theta_x(z), \theta_v(z)) \in \mathbb{R}^d \times \mathbb{R}^d$ for any $z \in \mathbb{R}^d$.

Proof. Letting i be any index x or v, we have

$$\int_{\mathbb{R}^{2d}} (z_i + \theta_i(z))\, d\mu_0(z) = \int_{\mathbb{R}^{2d}} w_i\, d\mu_1(w)$$

by measure image relation, with

$$\int_{\mathbb{R}^{2d}} z_i\, d\mu_0(z) = \int_{\mathbb{R}^{2d}} w_i\, d\mu_1(w)$$

by assumption on the moments, so that

$$\int_{\mathbb{R}^{2d}} \theta_i(z)\, d\mu_0(z) = 0.$$

In particular, expanding the square,

$$\begin{aligned}\iint_{\mathbb{R}^{4d}} |\theta_i(z) - \theta_i(w)|^2\, d\mu_0(z)\, d\mu_0(w) &= 2\int_{\mathbb{R}^{2d}} |\theta_i(z)|^2\, d\mu_0(z) - 2\Big|\int_{\mathbb{R}^{2d}} \theta_i(z)\, d\mu_0(z)\Big|^2 \\ &= 2\int_{\mathbb{R}^{2d}} |\theta_x(z)|^2\, d\mu_0(z)\end{aligned}$$

which concludes the argument. □

For general initial data, Theorem 1.3 can be made more precise for large t as in

Theorem 1.5. *Assume that V is twice differentiable, with Hessian matrix such that $-L\,I \leq D^2_x V(x) \leq L\,I$ for some constant L and any x in $\mathbb{R}^d$. If f and g are solutions to* (1.1) *with initial data f_0 and g_0 in $\mathcal{P}_2(\mathbb{R}^{2d})$, then*

$$W_2(f_t, g_t) \leq \inf\Big(2\,e^{\frac{1+L}{2}\,t} + \sqrt{2(1+t^2)}, e^{\left(\frac{1}{2}+L\right)t}\Big)\, W_2(f_0, g_0)$$

for any $t \geq 0$.

Let us note that the extra terms that appear when the initial data do not have the same means can certainly be improved, and why not removed. Note however that difficulties arising from the behaviour of the center of mass have been encountered in some other contexts, which has led to fix this center of mass : see for instance the works [36, 37] on the asymptotic behaviour of a granular media equation or [75] on its particle approximation, and [34] on the porous medium equation.

Proof. In view of the first statement in Theorem 1.3 we just prove the first bound ; for this we reduce the problem to that of the second case in Theorem 1.3 by translating f_0 into a measure h_0 with same center of mass and mean velocity as g_0.

More precisely we let

$$\mathcal{X} = \int x\, df_0(x,v) - \int x\, dg_0(x,v) \quad \text{and} \quad \mathcal{V} = \int v\, df_0(x,v) - \int v\, dg_0(x,v)$$

be the difference of the centers of mass and mean velocities. Given $a \in \mathbb{R}^{2d}$ we also let τ_a stand for the translation map $z \mapsto z - a$ on $\mathbb{R}^{2d}$. In this notation we let $h_0 = \tau_{(\mathcal{X},\mathcal{V})} \sharp f_0$ have same means as g_0. Then $h : t \mapsto h_t = \tau_{(\mathcal{X}+t\,\mathcal{V},\mathcal{V})} \sharp f_t$ is solution to the Vlasov equation with initial datum h_0

Indeed, if ϕ is a $\mathcal{C}^\infty$ function on $[0,+\infty[\times\mathbb{R}^{2d}$ with compact support, then

$$\int \phi_t(x,v)\, dh_t(x,v) - \int \phi_0(x,v)\, dh_0(x,v) = \int \psi_t(x,v)\, df_t(x,v) - \int \psi_0(x,v)\, df_0(x,v)$$

where ψ is the $\mathcal{C}^\infty$ function on $[0,+\infty[\times\mathbb{R}^{2d}$ with compact support defined by $\psi_t(x,v) = \phi_t \circ \tau_{(\mathcal{X}+t\,\mathcal{V},\mathcal{V})}(x,v) = \phi_t(x - \mathcal{X} - t\,\mathcal{V}, v - \mathcal{V})$. But f_t is a solution to the Vlasov equation with initial datum f_0, so the latter expression is equal to

$$-\int_0^t \int \Big(\frac{\partial \psi_s}{\partial s} + v \cdot \nabla_x \psi_s - \nabla_v \psi_s \cdot \nabla_x V * \rho[f_s]\Big)(x,v)\, df_s(x,v)\, ds. \tag{1.10}$$

Noting in particular that

$$\frac{\partial \psi_s}{\partial s}(x,v) = \frac{\partial \phi_s}{\partial s}(x - \mathcal{X} - s\,\mathcal{V}, v - \mathcal{V}) - \mathcal{V} \cdot \nabla_x \phi_s(x - \mathcal{X} - t\,\mathcal{V}, v - \mathcal{V})$$

and

$$\nabla_x V * \rho[f_s](x,v) = \nabla_x V * \rho[h_s](x - \mathcal{X} - s\,\mathcal{V})$$

since $h_s = \tau_{(\mathcal{X}+s\,\mathcal{V},\mathcal{V})}\sharp f_s$, it follows that (1.10) is in turn equal to

$$-\int_0^t \int \Big(\frac{\partial \phi_s}{\partial s}(x,v) + v \cdot \nabla_x \phi_s(x,v) - \nabla_v \phi_s(x,v) \cdot \nabla_x V * \rho[h_s](x)\Big)\, dh_s(x,v)\, ds.$$

This means that indeed h is a solution to the Vlasov equation with initial datum h_0.

Since h_0 and g_0 have same center of mass and mean velocity, Theorem 1.3 ensures that the distance between their respective solutions at time t satisfies

$$W_2(h_t, g_t) \leq e^{\frac{1+L}{2}\,t}\, W_2(h_0, g_0).$$

We finally get an estimate on f_t and g_t from this bound.

First of all
$$W_2(f_t, h_t) = \big(|\mathcal{X} + t\,\mathcal{V}|^2 + |\mathcal{V}|^2\big)^{1/2}$$
since h_t is obtained by translating f_t according to the vector $(\mathcal{X} + t\,\mathcal{V}, \mathcal{V})$. This leads us to bound $\big(|\mathcal{X} + t\,\mathcal{V}|^2 + |\mathcal{V}|^2\big)^{1/2}$ in terms of f_0 and g_0.

In $t = 0$, for any unit vector (e_x, e_v) in $\mathbb{R}^{2d}$ we have

$$<(\mathcal{X},\mathcal{V}),(e_x,e_v)> = \mathcal{X}\cdot e_x + \mathcal{V}\cdot e_v = \int (x\cdot e_x + v\cdot e_v)\,d(f_0 - g_0)(x,v)$$

by definition of $\mathcal{X}$ and $\mathcal{V}$. But the map $(x,v) \mapsto x\cdot e_x + v\cdot e_v$ is 1-Lipschitz on $\mathbb{R}^{2d}$, so Kantorovich-Rubinstein dual formulation ensures that

$$<(\mathcal{X},\mathcal{V}),(e_x,e_v)> \leq W_1(f_0, g_0).$$

Since this holds for any unit vector (e_x, e_v), and since $W_1 \leq W_2$, we finally obtain

$$\big(|\mathcal{X}|^2 + |\mathcal{V}|^2\big)^{1/2} \leq W_2(f_0, g_0).$$

In other words, $W_2(h_0, f_0) \leq W_2(f_0, g_0)$ and $W_2(h_0, g_0) \leq 2\,W_2(f_0, g_0)$ by triangular inequality.

Then, for general $t \geq 0$, the map $(x,v) \mapsto (x + t\,v)\cdot e_x + v\cdot e_v$ is $(2(1+t^2)^{1/2}$-Lipschitz, so in the same way we obtain

$$W_2(f_t, h_t) = \big(|\mathcal{X} + t\,\mathcal{V}|^2 + |\mathcal{V}|^2\big)^{1/2} \leq (2(1+t^2))^{1/2}\, W_2(f_0, g_0).$$

Collecting all terms we conclude by triangular inequality that

$$\begin{aligned} W_2(f_t, g_t) \leq W_2(f_t, h_t) + W_2(h_t, g_t) &\leq (2(1+t^2))^{1/2}\, W_2(f_0, g_0) + e^{\frac{1+L}{2}\,t}\, W_2(h_0, g_0) \\ &\leq \big(2e^{\frac{1+L}{2}\,t} + (2(1+t^2))^{1/2}\big)\, W_2(f_0, g_0). \end{aligned}$$

This concludes the argument of Theorem 1.5. $\square$

Chapitre 2

Approximation particulaire des équations d'Euler incompressibles dans le plan en formulation vorticité

Dans ce chapitre nous nous intéressons à une méthode particulaire d'approximation des équations d'Euler incompressibles en dimension deux d'espace et en formulation vorticité, et montrons comment améliorer l'ordre de grandeur (en temps) de constantes apparaissant dans cette méthode en utilisant les distances de Wasserstein.

Introduction

Let us consider the two dimensional incompressible Euler equations

$$\begin{cases} \dfrac{\partial u}{\partial t} + (u \cdot \nabla)\, u &= -\nabla p \\ \nabla \cdot u &= 0 \end{cases} \qquad t > 0, \quad x \in \mathbb{R}^2 \tag{2.1}$$

where $u_t : x \mapsto u(t,x) \in \mathbb{R}^2$ is the *velocity field* of a fluid at time t, and $p_t : x \mapsto p(t,x)$ is called the *pressure field* at time t.

Here $a \cdot b$ denotes the scalar product of two vectors a and b in $\mathbb{R}^2$ and ∇ denotes the gradient operator ; in a consistent way $\nabla \cdot$ stands for the divergence operator, and $(u_t \cdot \nabla)\, u_t$ is the vector with components $\left(u_t^1 \dfrac{\partial u_t^i}{\partial x_1} + u_t^2 \dfrac{\partial u_t^i}{\partial x_2}\right)$ for $i = 1, 2$, if $u_t = (u_t^1, u_t^2)$ and $x = (x_1, x_2)$ in $\mathbb{R}^2$.

One can consider the *vorticity* ω_t of the fluid at time t defined as the rotational of the velocity field u_t, namely by

$$\omega_t = \frac{\partial u_t^2}{\partial x_1} - \frac{\partial u_t^1}{\partial x_2}.$$

The vorticity field ω_t gives a measure of how the fluid is rotating, and by (2.1) is transported by the velocity field u_t in the sense that $\omega : t \mapsto \omega_t$ satisfies the equation

$$\frac{\partial \omega}{\partial t} + u \cdot \nabla \omega = 0, \qquad t > 0, \quad x \in \mathbb{R}^2. \tag{2.2}$$

Formally the velocity u_t can be recovered from the vorticity ω_t in the following way (for this and for any other topics on the Euler equations we refer to [76] for instance). The incompressibility condition $\nabla \cdot u_t = 0$ is sufficient for the existence of a so-called stream function Ψ_t such that

$$u_t = \nabla^{\perp} \Psi_t$$

with the general notation $X^{\perp} = (X_2, -X_1)$ if $X = (X_1, X_2)$. Then Ψ_t is linked with the vorticity ω_t by the Poisson equation

$$\Delta \Psi_t = -\omega_t$$

where Δ is the Laplace operator in $\mathbb{R}^2$. This equation can be solved in $\mathbb{R}^2$ as $\Psi_t = G * \omega_t$ where G is the Green kernel

$$G(x) = -\frac{1}{2\pi} \ln |x|.$$

In particular

$$u_t = \nabla^{\perp} \Psi_t = K * \omega_t \tag{2.3}$$

where

$$K(x) = \nabla^{\perp} G(x) = -\frac{1}{2\pi} \frac{x^{\perp}}{|x|^2}.$$

Let us note that $K(x-y)$ is the velocity field in x generated by a charge of intensity one fixed in y. Let us also note that (2.3) makes sense if, for instance, ω_t belongs to $L^1(\mathbb{R}^2) \cap L^{\infty}(\mathbb{R}^2)$.

In particular (2.2) writes only in terms of the vorticity as

$$\frac{\partial \omega}{\partial t} + K * \omega \cdot \nabla \omega = 0, \qquad t > 0, \quad x \in \mathbb{R}^2.$$

In the Euler equations (2.1) the unknown quantity is the velocity field u_t. This means that we fix x and we look at the velocity of a particle that at time t passes through x : this is the Eulerian point of view.

Following some other point of view, one can label some particle initially positionned at some point x, and follow its trajectory $(\Phi_t(x))_{t \geq 0}$ as time goes ; this is the Lagrangian point of view, which is related to the Eulerian one by noting that u_t and Φ_t are linked by

$$\frac{d\Phi_t(x)}{dt} = u_t(\Phi_t(x)). \tag{2.4}$$

In particular, if one knows all the trajectories $(\Phi_t(x))_{t \geq 0}$ of the fluid particles initially at $\Phi_0(x) = x$, then one can recover the velocity field by differentiation. Conversely, knowing the velocity field, one can obtain the trajectories of the fluid particles by solving (2.4) from the initial datum $\Phi_0(x) = x$.

In this notation, (2.2) means that the vorticity is transported by the flow Φ_t. More precisely, as a measure, ω_t is the image measure $\Phi_t \sharp \omega_0$ of ω_0 by the map $\Phi_t : \mathbb{R}^2 \to \mathbb{R}^2$. The incompressibility condition imposed on the velocity u_t in (2.1) ensures that Φ_t is inversible with constant Jacobian determinant equal to 1, so that

$$\omega_t \; = \; \omega_0 \circ \Phi_t^{-1}$$

as the level of the vorticity densities.

To sum up the Euler equations (2.1) for some initial datum u_0 on $\mathbb{R}^2$ are formally equivalent to the system

$$\left\{ \begin{array}{rcll} \dfrac{d\Phi_t(x)}{dt} & = & u_t(\Phi_t(x)), & t > 0, \quad x \in \mathbb{R}^2 \\ \Phi_0(x) & = & x, & x \in \mathbb{R}^2 \\ \omega_t(x) & = & \omega_0(\Phi_t^{-1}(x)), & t > 0, \quad x \in \mathbb{R}^2 \\ u_t(x) & = & K * \omega_t(x), & t > 0, \quad x \in \mathbb{R}^2 \end{array} \right. \tag{2.5}$$

for some initial datum ω_0 on $\mathbb{R}^2$.

It is precisely this system we now consider.

Assuming that the initial vorticity ω_0 belongs to $L^1(\mathbb{R}^2) \cap L^\infty(\mathbb{R}^2)$, there exists a unique triple (Φ, ω, u) solution to (2.5), where ω belongs to $L^\infty([0, +\infty[\times \mathbb{R}^2) \cap L^\infty([0, +\infty[, L^1(\mathbb{R}^2))$ with $\|\omega_t\|_{L^p(\mathbb{R}^2)} = \|\omega_0\|_{L^p(\mathbb{R}^2)}$ for any $t \geq 0$ and $p \geq 1$ (see for instance [76, Exercise 2.9]).

From now on we shall assume that ω_0 is nonnegative with unit integral, and can be seen as a probability measure with bounded density (by identifying an absolutely continous measure to its density). Then the same holds for ω_t by the general properties mentionned above.

Let us note here that the relations $\|\omega_t\|_{L^p(\mathbb{R}^2)} = \|\omega_0\|_{L^p(\mathbb{R}^2)}$ are consequences of the fact that ω_t is the image measure of ω_0 by the measure-preserving map Φ_t.

2.1 Particle approximation

These equations have been studied from diverse points of view, but here we want to focus on some particle approximation of (2.5), which we now describe.

On one hand we let ω_0 be a probability measure on $\mathbb{R}^2$ with bounded density, evolving into ω_t according to equations (2.5).

On the other hand we let $X_0^1, \ldots, X_0^N$ be N points in $\mathbb{R}^2$ and $a_1, \ldots, a_N$ be N nonnegative numbers satisfying the normalization condition $\displaystyle\sum_{i=1}^N a_i = 1$. We let also K_ε be a divergence free $\mathcal{C}^\infty$ $\mathbb{R}^2$-valued function on $\mathbb{R}^2$, with $K_\varepsilon(0) = 0$ and equal to K but at short distances, namely such that $K_\varepsilon(z) = K(z)$ if $|z| > \varepsilon$, where ε is a given positive number. Let also $c_\varepsilon = \max(2\, B_\varepsilon, L_\varepsilon)$ where K_ε is bounded by B_ε and has Lipschitz seminorm bounded by L_ε.

Then we let the X_0^i's evolve according to

$$\begin{cases} \dfrac{dX_t^{i,\varepsilon}}{dt} = \displaystyle\sum_{j=1}^{N} a_j\, K_\varepsilon(X_t^{i,\varepsilon} - X_t^{j,\varepsilon}) \\ X_0^{i,\varepsilon} = X_0^i \end{cases} \qquad 1 \leq i \leq N.$$

By assumption on K_ε (and properties on K) this system of differential equations admits a unique solution, whereas it may not have if we had not replaced the genuine singular kernel K by K_ε (even taking a sum on $j \neq i$ in the right hand side).

Let also

$$\omega_t^{N,\varepsilon} = \sum_{i=1}^{N} a_i\, \delta_{X_t^{i,\varepsilon}}$$

be the vorticity profile made of the N vortices $a_i\, \delta_{X_t^{i,\varepsilon}}$ with intensity a_i (in some other context it would be called the empirical measure of the N particles $X_t^{i,\varepsilon}$ at time t with the masses a_i), and $\omega_0^{N,\varepsilon} = \omega_0^N$ initially.

Then the issue of the particle approximation in this model is the following : how good ω_0^N must be an approximation of ω_0, and how do we have to choose the kernel K_ε to be sure that $\omega_t^{N,\varepsilon}$ is a good approximation of ω_t at some later time t?

This problem is stated in a more precise form in [76, Section 5] by means of the Wasserstein distance $W_{1,d}$ of order 1 defined on the space $\mathcal{P}(\mathbb{R}^2)$ of probability measures on $\mathbb{R}^2$ from the distance d defined on $\mathbb{R}^2$ by

$$d(x,y) = \begin{cases} |x-y| & \text{if} \quad |x-y| < 1 \\ 1 & \text{otherwise} \end{cases}$$

where $|\,.\,|$ is the Euclidean norm on $\mathbb{R}^2$. The distance d is equivalent to the Euclidean distance on $\mathbb{R}^2$, and by the Kantorovich-Rubinstein dual formulation

$$W_{1,d}(\mu,\nu) = \sup\Big\{ \int_{\mathbb{R}^2} f(x)\, d\mu(x) - \int_{\mathbb{R}^2} f(x)\, d\mu(x);\ |f(y) - f(x)| \leq d(x,y) \text{ for all } x, y \Big\},$$

one can prove the relation

$$W_{1,d} \leq d_{BL} \leq 2\, W_{1,d}$$

where d_{BL} is the dual-bounded Lipschitz distance defined by

$$d_{BL}(\mu,\nu) = \sup\Big\{ \int_{\mathbb{R}^2} f(x)\, d\mu(x) - \int_{\mathbb{R}^2} f(x)\, d\mu(x);\ |f(y) - f(x)| \leq |x-y|,\ |f(x)| \leq 1 \text{ for all } x, y \Big\}.$$

Since d_{BL} metrizes the (narrow) weak topology on $\mathcal{P}(\mathbb{R}^2)$ as mentionned in the previous chapter, it follows that so does the distance $W_{1,d}$.

Then C. Marchioro and M. Pulvirenti [76] obtain the following result :

Theorem 2.1 (cf. [76]). *Assume that the kernel K_ε is bounded by some constant B_ε and Lipschitz with Lipschitz seminorm L_ε. Then, in the above notation, for any $T \geq 0$ there holds*

$$\sup_{0 \leq t \leq T} W_{1,d}(\omega_t, \omega_t^{N,\varepsilon(N)}) \longrightarrow 0 \quad as \quad N \longrightarrow +\infty$$

for any sequence $\varepsilon(N)$ such that

$$\exp\big(c_{\varepsilon(N)}(T + e^{c_{\varepsilon(N)}\,T})\big) W_{1,d}(\omega_0, \omega_0^N) \longrightarrow 0 \quad as \quad N \longrightarrow +\infty$$

where $c_\varepsilon = \max(2\,B_\varepsilon, L_\varepsilon)$.

As pointed out in [76], the result can be considerably improved under some smoothness assumption on the initial data and a better choice of K_ε.

In the next section we simply show how to improve the imposed condition by an exponential factor in a simple way, by just making some other choice of distance between the involved probability measures.

2.2 Improving the rate of convergence

Let us start by giving the sketch of the proof of Theorem 2.1 as given in [76].

The idea is to introduce a measure ω_t^ε which is far neither from ω_t nor from $\omega_t^{N,\varepsilon}$. For this purpose we let $\omega^\varepsilon : t \mapsto \omega_t^\varepsilon$ be the solution in the sense of distributions of the (regularized Euler) equation

$$\frac{\partial \omega^\varepsilon}{\partial t} + K_\varepsilon * \omega^\varepsilon \cdot \nabla \omega^\varepsilon = 0, \qquad t > 0, \quad x \in \mathbb{R}^2 \tag{2.6}$$

with initial datum $\omega_0^\varepsilon = \omega_0$. Such a solution exists and is unique, and satisfies the property

$$\sup_{0 \leq t \leq T} W_{1,d}(\omega_t^\varepsilon, \omega_t) \longrightarrow 0 \quad \text{as} \quad \varepsilon \longrightarrow 0 \tag{2.7}$$

where ω_t has been defined as the solution to the (true) Euler equation (2.5) with the same initial datum ω_0.

But $\omega^{N,\varepsilon}$ also is a solution to (2.6), and (2.6) satisfies the following general stability property : if μ^i for $i = 1, 2$ are two solutions to (2.6) with respective initial data μ_0^i in $\mathcal{P}(\mathbb{R}^2)$, then we have

$$W_{1,d}(\mu_t^1, \mu_t^2) \leq \exp\big(c_\varepsilon(t + e^{c_\varepsilon\,t})\big) W_{1,d}(\mu_0^1, \mu_0^2) \tag{2.8}$$

for any $t \geq 0$. In particular

$$W_{1,d}(\omega_t^\varepsilon, \omega_t^{N,\varepsilon}) \leq \exp\big(c_\varepsilon(t + e^{c_\varepsilon\,t})\big) W_{1,d}(\omega_0, \omega_0^N)$$

and by triangular inequality

$$\begin{aligned}
\sup_{0 \leq t \leq T} W_{1,d}(\omega_t, \omega_t^{N,\varepsilon}) &\leq \sup_{0 \leq t \leq T} W_{1,d}(\omega_t, \omega_t^\varepsilon) + \sup_{0 \leq t \leq T} W_{1,d}(\omega_t^\varepsilon, \omega_t^{N,\varepsilon}) \\
&\leq \sup_{0 \leq t \leq T} W_{1,d}(\omega_t, \omega_t^\varepsilon) + \exp\big(c_\varepsilon(T + e^{c_\varepsilon\,T})\big) W_{1,d}(\omega_0, \omega_0^N)
\end{aligned}$$

for any $T \geq 0$, which concludes the argument of Theorem 2.1 by (2.7). $\square$

We note that, as for the Vlasov equation discussed in the previous chapter, the convergence of the particle approximation is largely based on a (quantitative) stability result on a certain partial differential equation. In particular the accuracy of the method depends on how the constant that appears in the bound (2.8) grows in ε and t.

It is precisely this estimate (2.8) we now aim at improving.

For this purpose we consider again the Wasserstein distances W_p defined from the usual Euclidean distance $|x-y|$ on $\mathbb{R}^2$, on the set $\mathcal{P}_p(\mathbb{R}^2)$ of probability measures with finite moment of order p for the Euclidean norm.

Let us recall that we have let K_ε be a divergence free $\mathcal{C}^\infty$ $\mathbb{R}^2$-valued function on $\mathbb{R}^2$, with $K_\varepsilon(0)=0$ and such that $K_\varepsilon(z)=K(z)$ if $|z|>\varepsilon$. We have also let $c_\varepsilon = \max(2\,B_\varepsilon, L_\varepsilon)$ where K_ε is bounded by B_ε and has Lipschitz seminorm bounded by L_ε.

Any two solutions ω^1 and ω^2 to the (regularized Euler) equation

$$\frac{\partial \omega}{\partial t} + K_\varepsilon * \omega \cdot \nabla \omega = 0, \qquad t>0, \quad x \in \mathbb{R}^2 \tag{2.9}$$

for some initial data ω_0^1 and ω_0^2 in $\mathcal{P}(\mathbb{R}^2)$ have been proven in [76] to satisfy the relation

$$W_{1,d}(\omega_t^1, \omega_t^2) \leq \exp\big(c_\varepsilon(t+e^{c_\varepsilon\, t})\big) W_{1,d}(\omega_0^1, \omega_0^2)$$

for any $t \geq 0$.

We now improve this bound by an exponential factor, first using the W_1 distance :

Theorem 2.2. *Assume that the kernel K_ε is Lipschitz with Lipschitz seminorm L_ε. If ω^1 and ω^2 are solutions to (2.9) with initial data ω_0^1 and ω_0^2 in $\mathcal{P}_1(\mathbb{R}^{2d})$, then*

$$W_1(\omega_t^1, \omega_t^2) \leq e^{2L_\varepsilon\, t}\, W_1(\omega_0^1, \omega_0^2)$$

for any $t \geq 0$.

Proof. We adapt to our case the method developed in [85] or [104] for the Vlasov equation, and used also to prove Theorem 1.2 in the previous chapter.

That ω_t^i has finite moment, hence belongs to $\mathcal{P}_1(\mathbb{R}^2)$, can be proven by an a priori estimate on the quantity $\displaystyle\int_{\mathbb{R}^2} <x>\, d\omega_i^t(x)$, where $<x>= (1+|x|^2)^{1/2}$, as done in detail in the previous chapter.

Then, to ω_t^i for $i=1,2$ we associate the corresponding flow Φ_t^i given for $x \in \mathbb{R}^2$ by

$$\begin{cases} \dfrac{d\Phi_t^i(x)}{dt} &= K_\varepsilon * \omega_t^i(\Phi_t^i(x)), \\ \Phi_0^i(x) &= x. \end{cases}$$

Note that existence and uniqueness to these equations result from the Lipschitz property of K_ε and hence of $K_\varepsilon * \omega_t^i$, and that ω_t^i is the image measure $\Phi_t^i \sharp \omega_t^i$ of ω_0^i by the measure-preserving map Φ_t (since K_ε is divergence free).

By definition of the W_1 distance we shall have to consider joint measures on $\mathbb{R}^2 \times \mathbb{R}^2$ with marginals ω_t^1 and ω_t^2. Some of them can be built in the following way : let π_0 be any probability measure on $\mathbb{R}^2 \times \mathbb{R}^2$ with marginals ω_0^1 and ω_0^2, and $\Phi_t(x,y) = (\Phi_t^1(x), \Phi_t^2(y))$; then the measure $\pi_t = \Phi_t \sharp \pi_0$ has marginals ω_t^1 and ω_t^2.

Indeed, if P_1 and P_2 are the projections defined by $P_1(x,y) = x$ and $P_2(x,y) = y$, the identities

$$P_i \circ \Phi_t = \Phi_t^i \circ P_i$$

hold for $i = 1, 2$, so that for any continuous bounded functions a and b on $\mathbb{R}^2$ we have

$$\begin{aligned}
\int_{\mathbb{R}^2\times\mathbb{R}^2} (a(x)+b(y))\; d\pi_t(x,y) &= \int_{\mathbb{R}^2\times\mathbb{R}^2} (a\circ P_1\circ\Phi_t(x,y) \;+\; b\circ P_2\circ\Phi_t(x,y))\; d\pi_0(x,y) \\
&= \int_{\mathbb{R}^2\times\mathbb{R}^2} \big(a\circ\Phi_t^1\circ P_1(x,y) \;+\; b\circ\Phi_t^2\circ P_2(x,y)\big)\; d\pi_0(x,y) \\
&= \int_{\mathbb{R}^2\times\mathbb{R}^2} \big(a\circ\Phi_t^1(x) \;+\; b\circ\Phi_t^2(y)\big)\; d\pi_0(x,y) \\
&= \int_{\mathbb{R}^2} a\circ\Phi_t^1(x)\, d\omega_0^1(x) \;+\; \int_{\mathbb{R}^2} b\circ\Phi_t^2(y)\, d\omega_0^2(y) \\
&= \int_{\mathbb{R}^2} a(x)\, d\omega_t^1(x) \;+\; \int_{\mathbb{R}^2} b(y)\, d\omega_t^2(y).
\end{aligned}$$

This means that indeed π_t has marginals ω_t^1 and ω_t^2.

Consequently

$$W_1(\omega_t^1,\omega_t^2) \;\le \int_{\mathbb{R}^2\times\mathbb{R}^2} |x-y|\, d\pi_t(x,y) \;=\; \int_{\mathbb{R}^2\times\mathbb{R}^2} |\Phi_t^1(x)-\Phi_t^2(y)|\, d\pi_0(x,y) \;\le\; I_1(t)+I_2(t) \tag{2.10}$$

where

$$I_1(t) = \int_{\mathbb{R}^2\times\mathbb{R}^2} |\Phi_t^1(x)-\Phi_t^2(x)|\, d\pi_0(x,y) = \int_{\mathbb{R}^2} |\Phi_t^1(x)-\Phi_t^2(x)|\, d\omega_0^1(x)$$

corresponds to particles having the same initial position, but moving according to different flows, and

$$I_2(t) \;=\; \int_{\mathbb{R}^2\times\mathbb{R}^2} |\Phi_t^2(x)-\Phi_t^2(y)|\, d\pi_0(x,y)$$

corresponds to particles having diverse initial positions, but moving according to the same flow.

First of all, as regards the I_1 term,

$$\frac{dI_1(t)}{dt} \;=\; \int_{\mathbb{R}^2} \frac{d}{dt}\big|\Phi_t^1(x)-\Phi_t^2(x)\big|\, d\omega_0^1(x)$$

where

$$\begin{aligned}
\frac{d}{dt}\big|\Phi_t^1(x)-\Phi_t^2(x)\big| \;&\le\; \left|\frac{d}{dt}\{\Phi_t^1(x)-\Phi_t^2(x)\}\right| = |K_\varepsilon*\omega_t^1(\Phi_t^1(x)) - K_\varepsilon*\omega_t^2(\Phi_t^2(x))| \\
&\le\; |K_\varepsilon*\omega_t^1(\Phi_t^1(x)) - K_\varepsilon*\omega_t^1(\Phi_t^2(x))| \;+\; |K_\varepsilon*\omega_t^1(\Phi_t^2(x)) - K_\varepsilon*\omega_t^2(\Phi_t^2(x))|.
\end{aligned}$$

But the first term is bounded by $L_\varepsilon\,|\Phi_t^1(x) - \Phi_t^2(x)|$ since K_ε, and hence $K_\varepsilon * \omega_t^1$, is L_ε-Lipschitz, and for the same reason the second term is bounded by $L_\varepsilon\, W_1(\omega_t^1, \omega_t^2)$ by the Kantorovich-Rubinstein dual formulation of W_1.

Hence

$$\frac{dI_1(t)}{dt} \,\leq\, L_\varepsilon\, I_1(t) \,+\, L_\varepsilon\, W_1(\omega_t^1, \omega_t^2),$$

by integrating with respect to the measure ω_0^1, so that

$$I_1(t)\, e^{-L_\varepsilon\, t} \,\leq\, L_\varepsilon \int_0^t W_1(\omega_s^1, \omega_s^2)\, e^{-L_\varepsilon\, s}\, ds \tag{2.11}$$

since $I_1(0) = 0$.

Then the I_2 term satisfies the inequality

$$\frac{dI_2(t)}{dt} \,\leq\, L_\varepsilon\, I_2(t)$$

since in the same way

$$\frac{d}{dt}|\Phi_t^2(x) - \Phi_t^2(y)| \,\leq\, |K_\varepsilon * \omega_t^2(\Phi_t^2(x)) - K_\varepsilon * \omega_t^2(\Phi_t^2(y))| \,\leq\, L_\varepsilon\, |\Phi_t^2(x) - \Phi_t^2(y)|.$$

Thus

$$I_2(t)e^{-L_\varepsilon\, t} \,\leq\, I_2(0) \,=\, \int_{\mathbb{R}^2\times\mathbb{R}^2} |\Phi_0^1(x) - \Phi_0^2(y)|\, d\pi_0(x,y) \,=\, \int_{\mathbb{R}^2\times\mathbb{R}^2} |x-y|\, d\pi_0(x,y). \tag{2.12}$$

Collecting (2.10), (2.11) and (2.12) we obtain the inequality

$$W_1(\omega_t^1, \omega_t^2)e^{-L_\varepsilon\, t} \,\leq\, (I_1 + I_2)(t)\, e^{-L_\varepsilon\, t} \,\leq\, I_2(0) + L_\varepsilon \int_0^t W_1(\omega_s^1, \omega_s^2)\, e^{-L_\varepsilon\, s}\, ds,$$

which by Gronwall lemma leads to

$$W_1(\omega_t^1, \omega_t^2) \,\leq\, I_2(0)\, e^{2L_\varepsilon\, t} \,=\, \int_{\mathbb{R}^2\times\mathbb{R}^2} |x-y|\, d\pi_0(x,y)\, e^{2L_\varepsilon\, t}.$$

Taking the infimum over all π_0 with marginals ω_0^1 and ω_0^2 concludes the argument. □

We now give an analogous result in W_2 distance :

Theorem 2.3. *Assume that the kernel K_ε is odd and has derivative $\nabla_x K_\varepsilon$ with Euclidean matrix norm uniformly bounded by some constant L_ε. If ω^1 and ω^2 are solutions to (2.9) with initial data f_0 and g_0 in $\mathcal{P}_2(\mathbb{R}^{2d})$, then*

$$W_2(f_t, g_t) \leq e^{2L_\varepsilon\, t}\, W_2(f_0, g_0)$$

for any $t \geq 0$.

If moreover ω_0^1 and ω_0^2 have same center of mass, then

$$W_2(\omega_t^1, \omega_t^2) \leq e^{L_\varepsilon\, t}\, W_2(\omega_0^1, \omega_0^2)$$

for any $t \geq 0$.

By center of mass of a measure μ on $\mathbb{R}^2$ we mean the moment $\displaystyle\int_{\mathbb{R}^2} x\, d\mu(x)$. As in the example of the Vlasov equation we have improved the constant by passing from the simpler distance W_1 to the more complex W_2 ; note that the oddness condition on K_ε is not so strong since the original kernel K itself is odd.

Proof. We adapt the formal argument used in the previous chapter for the Vlasov equation (Theorem 1.3), to which we refer for further details.

We first note that, since K_ε is an odd function, the moment $\displaystyle\int_{\mathbb{R}^2} |x|^2\, d\omega_t^i(x)$ is conserved by the equation, so that ω_t^i belongs to $\mathcal{P}_2(\mathbb{R}^2)$ as soon as so does ω_0^i. This conservation property is actually shared by the original Euler equation (2.5).

Given $t \geq 0$, and using the same notation as in the previous chapter, we assume that $\mu_0 = \omega_t^1$ and $\mu_1 = \omega_t^2$ are absolutely continuous with respect to the Lebesgue measure, so that there exists an optimal (for W_2) map T between μ_0 and μ_1. Then we let $T^s = (1-s)\,T + s\, Id$ and $\mu_s = T^s \sharp \mu_0$ for $s \in [0,1]$ and consider the curve $(\mu_s)_{0 \leq s \leq 1}$ with endpoints μ_0 and μ_1. We also introduce the velocity field associated to this curve and given by

$$v_s(T^s(x)) = \frac{dT^s(x)}{ds} = T(x) - x.$$

On the other hand K_ε is divergence free, so equation (2.9) writes as the transport equation

$$\frac{\partial \omega_t}{\partial t} + \nabla \cdot (\omega_t\, K_\varepsilon * \omega_t) \;=\; 0.$$

Hence the squared distance $W_2^2(\omega_t^1, \omega_t^2)$ between the two solutions ω_t^1 and ω_t^2 is derivable with respect to time, with

$$\frac{1}{2}\frac{d}{dt} W_2^2(\omega_t^1, \omega_t^2) = \int_{\mathbb{R}^2} v_1(x) \cdot K_\varepsilon * \omega_t^2(x)\, d\omega_t^2(x) - \int_{\mathbb{R}^2} v_0(x) \cdot K_\varepsilon * \omega_t^1(x)\, d\omega_t^1(x) \tag{2.13}$$

that we write as

$$F(1) - F(0) = \int_0^1 F'(s)\, ds$$

where

$$F(s) = \int_{\mathbb{R}^2} v_s(x) \cdot K_\varepsilon * \mu_s(x)\, d\mu_s(x) = \int_{\mathbb{R}^2} \theta(x) \cdot K_\varepsilon * \mu_s(T^s(x))\, d\mu_0(x)$$

in the notation $\theta(x) = T(x) - x$.

But

$$K_\varepsilon * \mu_s(T^s(x)) = \int_{\mathbb{R}^2} K_\varepsilon\left(T^s(x) - y\right)\, d\mu_s(y) = \int_{\mathbb{R}^2} K_\varepsilon\left(x - y + s(\theta(x) - \theta(y))\right)\, d\mu_0(y),$$

so

$$\frac{d}{ds} K_\varepsilon * \mu_s(T^s(x)) = \int_{\mathbb{R}^2} \nabla_x K_\varepsilon\left(x - y + s(\theta(x) - \theta(y))\right)\, (\theta(x) - \theta(y))\, d\mu_0(y)$$

where $\nabla_x K_\varepsilon$ is the space derivative of the map $K_\varepsilon : \mathbb{R}^2 \to \mathbb{R}^2$. Since K_ε is assumed to be odd, $\nabla_x K_\varepsilon$ is symmetric, so that

$$F'(s) = \frac{1}{2} \iint_{\mathbb{R}^2\times\mathbb{R}^2} (\theta(x)-\theta(y)) \cdot \nabla_x K_\varepsilon \left(x-y+s(\theta(x)-\theta(y))\right) (\theta(x)-\theta(y))\, d\mu_0(x)\, d\mu_0(y).$$

Moreover $\nabla_x K_\varepsilon$ has Euclidean matrix norm uniformly bounded by L_ε, so

$$F'(s) \leq \frac{L_\varepsilon}{2} \iint_{\mathbb{R}^2\times\mathbb{R}^2} |\theta(x)-\theta(y)|^2\, d\mu_0(x)\, d\mu_0(y) \tag{2.14}$$

for any s dans $[0,1]$.

In the general case when ω_0^1 and ω_0^2 are not assumed to have same center of mass, the right-hand side in (2.14) is bounded by

$$2\, L_\varepsilon \int_{\mathbb{R}^2} |\theta(x)|^2\, d\mu_0(x) = 2\, L_\varepsilon\, W_2^2(\omega_t^1, \omega_t^2)$$

since $Id+\theta$ is an optimal map between $\mu_0 = \omega_t^1$ and $\mu_1 = \omega_t^2$. Hence, by (2.13) and (2.14),

$$\frac{1}{2}\frac{d}{dt} W_2^2(\omega_t^1, \omega_t^2) \leq 2\, L_\varepsilon\, W_2^2(\omega_t^1, \omega_t^2),$$

which ensures the first result in Theorem 2.3 by Gronwall's lemma.

In the specific case when ω_0^1 and ω_0^2 are assumed to have same center of mass, we first note that so do ω_t^1 and ω_t^2. Indeed, on one hand

$$\frac{d}{dt}\int_{\mathbb{R}^2} x\, d\omega_t^i(x) = \int_{\mathbb{R}^2} K_\varepsilon * \omega_t^i(x)\, d\omega_t^i(x) = \iint_{\mathbb{R}^2\times\mathbb{R}^2} K_\varepsilon(x-y)\, d\omega_t^i(x)\, d\omega_t^i(y),$$

and on the other hand K_ε is an odd function, so that this integral is equal to its opposite, hence to 0, which means that $\displaystyle\int_{\mathbb{R}^2} x\, d\omega_t^i(x)$ indeed is conserved.

Since moreover the map $Id+\theta$ transports μ_0 onto μ_1, Lemma 1.4, written with only one (space) variable, ensures that

$$\iint_{\mathbb{R}^2\times\mathbb{R}^2} |\theta(x)-\theta(y)|^2\, d\mu_0(x)\, d\mu_0(y) = 2\int_{\mathbb{R}^2} |\theta(x)|^2\, d\mu_0(x). \tag{2.15}$$

But again $Id+\theta$ is an optimal map between $\mu_0 = \omega_t^1$ and $\mu_1 = \omega_t^2$, so one can proceed as in the first case to obtain the bound

$$\frac{1}{2}\frac{d}{dt} W_2^2(\omega_t^1, \omega_t^2) \leq L_\varepsilon\, W_2^2(\omega_t^1, \omega_t^2).$$

from (2.13), (2.14) and (2.15), and conclude the argument of Theorem 2.3. □

As for the Vlasov equation, Theorem 2.3 for general initial data in $\mathcal{P}_2(\mathbb{R}^{2d})$ can be made more precise for large t as in

Theorem 2.4. *Assume that the kernel K_ε is odd and has derivative $\nabla_x K_\varepsilon$ with Euclidean matrix norm uniformly bounded by some constant L_ε. If $f\omega^1$ and ω^2 are solutions to* (2.9) *with initial data ω_0^1 and ω_0^2 in $\mathcal{P}_2(\mathbb{R}^{2d})$, then*

$$W_2(\omega_t^1, \omega_t^2) \leq \inf(2\, e^{L_\varepsilon\, t} + 1, e^{2\, L_\varepsilon t})\, W_2(\omega_0^1, \omega_0^2)$$

for any $t \geq 0$.

2.3 The Euler equations in a bounded domain

In the previous section we have proven a strong stability result for measure solutions to equation (2.9), which can be seen as a regularized version of the Euler equation in vorticity formulation.

Obtaining such a result for the genuine Euler equations (2.5) seems to be a much harder task, and in this section we aim at deriving a partial result in this direction.

We now let D be a simply connected and open set in $\mathbb{R}^2$, sufficiently regular so that one can define a normal n at each point of the boundary ∂D, and solve in a unique way the Poisson equation on D, with Dirichlet condition on ∂D : then we note G_D its fundamental solution and $K_D = \nabla^\perp G_D$.

As in the case of the whole space $\mathbb{R}^2$, the problem

$$\left\{ \begin{array}{rcll} \dfrac{\partial u}{\partial t} + (u \cdot \nabla)\, u & = & -\,\nabla p, & t > 0, \quad x \in D \\ \nabla \cdot u & = & 0, & t > 0, \quad x \in D \\ u.n & = & 0, & t > 0, \quad x \in \partial D \end{array} \right.$$

with initial datum u_0 on D can formally be written in the form

$$\left\{ \begin{array}{rcll} \dfrac{d\Phi_t(x)}{dt} & = & u_t(\Phi_t(x)), & t > 0, \quad x \in D \\ \Phi_0(x) & = & x, & x \in D \\ \omega_t(x) & = & \omega_0(\Phi_t^{-1}(x)), & t > 0, \quad x \in D \\ u_t(x) & = & K_D * \omega_t(x), & t > 0, \quad x \in D \end{array} \right. \tag{2.16}$$

for some initial datum ω_0 on D.

According to [76, Section 2.3], for any initial datum ω_0 in $L^\infty(D)$ there exists a unique triple (Φ, ω, u) solution to (2.16), where ω belongs to $L^\infty([0, +\infty[\times \mathbb{R}^2)$ and, for any $t \geq 0$ u_t is quasi-Lipschitz in the sense that

$$|u_t(x) - u_t(y)| \leq C\, \|\omega_0\|_{L^\infty}\, \varphi(|x - y|), \qquad x, y \in D \tag{2.17}$$

where the function φ is defined as

$$\varphi(r) = \left\{ \begin{array}{ll} r(1 - \ln r) & \text{if } r \leq 1 \\ 1 & \text{if } r > 1 \end{array} \right.$$

It can indeed be checked that the kernel K_D satisfies the bound

$$\int_D |K_D(x-y) - K_D(x'-y)|\, dy \;\leq\; C\,\varphi(|x-x'|) \tag{2.18}$$

for some constant C depending on D by its Lebesgue measure.

In the previous section when K_ε was a Lipschitz kernel, $K_\varepsilon * \omega_t$ also was Lipschitz, which we largely used in particular in connection with the Kantorovich-Rubinstein formulation. Now $u_t = K * \omega_t$ satisfies the weaker property (2.17), which will turn the exponential term obtained in the stability result Theorem 2.2 into a double-exponential term.

More precisely we prove

Theorem 2.5. *Let ω_0^1 and ω_0^2 be two probability measures on D with bounded density with respect to the Lebesgue measure, and such that $W_1(\omega_0^1, \omega_0^2) \leq 1$. If ω^1 and ω^2 are the solutions to the Euler equations* (2.16) *for the initial data ω_0^1 and ω_0^2, then*

$$W_1(\omega_t^1, \omega_t^2) \leq e^{1-\exp(-c\,t)}\, W_1(\omega_0^1, \omega_0^2)^{\exp(-c\,t)}$$

for any $t \leq \ln\big(1 - \ln W_1(\omega_0^1, \omega_0^2)\big)/c$, where $c = 3\,C\, \max\big(\|\omega_0^1\|_{L^\infty(D)}, \|\omega_0^2\|_{L^\infty(D)}\big)$ for some constant C depending on D (introduced in (2.18)*).*

Note that, D being bounded, any two probability measures on D belong to $\mathcal{P}_1(D)$ and are at most distant of the diameter of D in W_1 distance ; in particular the condition $W_1(\omega_0^1, \omega_0^2) \leq 1$ is satisfied for any initial data as soon as the diameter of D is not bigger than 1.

Proof. We use the same notation as in the proof of Theorem 2.2. In particular we let π_0 be some measure on $D \times D$ having marginals ω_0^1 and ω_0^2, and let $\Phi_t(x,y) = (\Phi_t^1(x), \Phi_t^2(y))$ where Φ_t^i is the flow associated to ω_t^i for $i = 1, 2$. Noting that $\pi_t = \Phi_t \sharp \pi_0$ has marginals ω_t^1 and ω_t^2, we write

$$W_1(\omega_t^1, \omega_t^2) \;\leq\; \int_{D\times D} |x-y|\, d\pi_t(x,y) \;=\; \int_{D\times D} |\Phi_t^1(x) - \Phi_t^2(y)|\, d\pi_0(x,y) \;\leq\; I_1(t) + I_2(t)$$

where

$$I_1(t) = \int_{D\times D} |\Phi_t^1(x) - \Phi_t^2(x)|\, d\pi_0(x,y) = \int_D |\Phi_t^1(x) - \Phi_t^2(x)|\, d\omega_0^1(x)$$

and

$$I_2(t) \;=\; \int_{D\times D} |\Phi_t^2(x) - \Phi_t^2(y)|\, d\pi_0(x,y).$$

First of all, as regards I_1 we have

$$\begin{aligned}
\frac{dI_1(t)}{dt} \;&\leq\; \int_D |K*\omega_t^1(\Phi_t^1(x)) - K*\omega_t^2(\Phi_t^2(x))|\, d\omega_0^1(x)\\
&\leq\; \int_D |K*\omega_t^1(\Phi_t^1(x)) - K*\omega_t^1(\Phi_t^2(x))|\, d\omega_0^1(x)\\
&\quad + \int_D |K*\omega_t^1(\Phi_t^2(x)) - K*\omega_t^2(\Phi_t^2(x))|\, d\omega_0^1(x).
\end{aligned} \tag{2.19}$$

Inequality (2.18) and then Jensen's inequality applied to the concave function φ ensure that the first term in the right hand side in (2.19) is bounded by

$$\begin{aligned} C\|\omega_t^1\|_{L^\infty(D)} &\int_D \varphi(|\Phi_t^1(x)-\Phi_t^2(x)|)\,d\omega_0^1(x) \\ &\le C\|\omega_t^1\|_{L^\infty(D)}\varphi\left(\int_D |\Phi_t^1(x)-\Phi_t^2(x)|\,d\omega_0^1(x)\right) = C\|\omega_t^1\|_{L^\infty(D)}\varphi(I_1(t)). \end{aligned} \tag{2.20}$$

To bound the second term one moreover uses the marginal properties of π_t and then the fact that the Jacobian of Φ_t^2 is 1 since the vector field $K*\omega_t^2$ is divergence-free. One successively gets

$$\begin{aligned} &\int_D |K*\omega_t^1(\Phi_t^2(x)) - K*\omega_t^2(\Phi_t^2(x))|\,d\omega_0^1(x) \\ &= \int_D \left|\int_{D\times D} \big[K(\Phi_t^2(x)-X) - K(\Phi_t^2(x)-Y)\big]\,d\pi_t(X,Y)\right| d\omega_0^1(x) \\ &\le \int_{D\times D} \left\{\int_D \big|K(\Phi_t^2(x)-X) - K(\Phi_t^2(x)-Y)\big|\;\|\omega_0^1\|_{L^\infty(D)}\,dx\right\} d\pi_t(X,Y) \\ &= \|\omega_0^1\|_{L^\infty(D)} \int_{D\times D}\left\{\int_{\Phi_t^2(D)} |K(y-X)-K(y-Y)|\;1\,dy\right\} d\pi_t(X,Y) \\ &\le C\,\|\omega_0^1\|_{L^\infty(D)} \int_{D\times D} \varphi(|X-Y|)\,d\pi_t(X,Y) \\ &\le C\,\|\omega_0^1\|_{L^\infty(D)}\varphi\left(\int_{D\times D} |\Phi_t^1(x)-\Phi_t^2(y)|\,d\pi_0(x,y)\right) \\ &\le C\,\|\omega_0^1\|_{L^\infty(D)}\varphi\left(I_1(t)+I_2(t)\right). \end{aligned}$$

But $\|\omega_t^1\|_{L^\infty(D)} = \|\omega_0^1\|_{L^\infty(D)}$ for any t by a general property of the Euler equation, so finally

$$\frac{dI_1}{dt}(t) \le C\,\|\omega_0^1\|_{L^\infty(D)}[\varphi(I_1(t)) + \varphi\left(I_1(t)+I_2(t)\right)].$$

As regards I_2, one repeats the argument leading to (2.20) to get the inequality

$$\frac{dI_2}{dt}(t) \le C\,\|\omega_t^2\|_{L^\infty(D)}\varphi(I_2(t)) = C\,\|\omega_0^2\|_{L^\infty(D)}\varphi(I_2(t))$$

Adding both terms leads to

$$\frac{d}{dt}(I_1+I_2)(t) \le C\,\max\big(\|\omega_0^1\|_{L^\infty(D)}, \|\omega_0^2\|_{L^\infty(D)}\big)\big(\varphi(I_1(t)) + \varphi(I_2(t)) + \varphi(I_1(t)+I_2(t))\big).$$

and then

$$\frac{d}{dt}(I_1+I_2)(t) \le c\,\varphi\left((I_1+I_2)(t)\right) \tag{2.21}$$

with $c = 3\,C\,\max\big(\|\omega_0^1\|_{L^\infty(D)}, \|\omega_0^2\|_{L^\infty(D)}\big)$ since φ is nondecreasing.

But the solution to the problem

$$\begin{cases} \dfrac{dy(t)}{dt} = c\,\varphi(y(t)) \\ y(0) \in]0,1], \end{cases}$$

which is nondecreasing since φ is nonnegative, is given by

$$y(t) = e^{1-\exp(-ct)}\, y(0)^{\exp(-ct)} \tag{2.22}$$

as far as $y(t) \in]0,1]$, that is, for all $t \leq \dfrac{1}{c} \ln\big(1 - \ln y(0)\big)$.

Let us now assume that initially $W_1(\omega_0^1, \omega_0^2) \leq 1$, and let π_0 be an optimal plan between ω_0^1 and ω_0^2. Then $(I_1 + I_2)(0)$ is equal to $W_1(\omega_0^1, \omega_0^2)$, and hence is not bigger than 1. By a comparison principle, (2.21) and (2.22) with $y(0) = (I_1 + I_2)(0)$ ensure that

$$(I_1 + I_2)(t) \leq e^{1-\exp(-ct)} \big((I_1 + I_2)(0)\big)^{\exp(-ct)}$$

for all $t \leq \dfrac{1}{c} \ln\Big(1 - \ln\big(I_1 + I_2)(0)\big)\Big)$. This concludes the argument since $(I_1 + I_2)(0) = W_1(\omega_0^1, \omega_0^2)$ and $W_1(\omega_t^1, \omega_t^2) \leq (I_1 + I_2)(t)$ at later times. $\square$

Chapitre 3

Métriques contractantes pour des lois de conservation scalaires

Ce chapitre est en grande partie une version précisée de l'article [21] écrit en collaboration avec Yann Brenier et Grégoire Loeper, et publié dans Journal of Hyperbolic Differential Equations ; un complément est apporté dans le dernier paragraphe.

Nous considérons des solutions entropiques croissantes de lois de conservation scalaires en une dimension d'espace, et montrons que les dérivées spatiales de telles solutions vérifient une propriété de contraction pour toutes les distances de Wasserstein. Ce résultat prolonge la propriété de contraction dans L^1 montrée par Kružkov. Nous traitons les lois de conservation scalaires visqueuses ou non, ainsi qu'une extension de ce résultat au cas de solutions de lois de conservation de fonctions de flux différents.

Introduction

Existence and uniqueness of solutions to scalar conservation laws in one space dimension have been established by Kružkov in the framework of entropy solutions (see [65] for instance), and among the properties satisfied by these solutions it is known that the L^1 norm between any two of them is a non-increasing function of time.

In this work we shall focus on a class of entropy solutions such that a certain distance between the space derivatives of any two such solutions be also nonincreasing in time. On this class of solutions this result extends the L^1 norm contraction property.

More precisely we consider as initial data nondecreasing functions on $\mathbb{R}$ with limits 0 and 1 at $-\infty$ and $+\infty$ respectively. These properties are preserved by the conservation law, and corresponding solutions have been shown in [30] to arise in some models of pressureless gases, obtained as a continuous limit of systems of sticky particles. Noticing that the distributional space derivative of these functions are probability measures, we may consider the Wasserstein distance between the space derivatives of any two such solutions, and we shall prove in this paper that this distance is a nonincreasing function of time, constant in the case of classical solutions.

3.1 Introduction to the results

Given a locally Lipschitz real-valued function f on $\mathbb{R}$, called a flux, we consider the scalar conservation law

$$\begin{cases} u_t + f(u)_x \,=\, 0\,, \qquad t>0,\ x\in\mathbb{R}, \\ u(0,.) \,=\, u^0, \end{cases} \tag{3.1}$$

with unknown $u = u(t,x) \in \mathbb{R}$ and initial datum $u^0 \in L^\infty(\mathbb{R})$.

Notation : In this work u_t is the derivative of u with respect to t (and not the function $u(t,.)$ as in the previous chapters), and u_x is the derivative of u with respect to x.

We shall consider solutions that are called entropy solutions (see [101] for instance) and are defined as follows : a function $u = u(t,x) \in L^\infty(]0,+\infty[\times\mathbb{R})$ is said to be an entropy solution of (3.1) on $[0,+\infty[\times\mathbb{R}$ if the entropy inequality

$$E(u)_t \,+\, F(u)_x \,\le\, 0 \tag{3.2}$$

holds in the sense of distributions for any convex Lipschitz function E on $\mathbb{R}$, and with associated flux F defined by

$$F(u) \,=\, \int_0^u f'(v)\,E'(v)\,dv. \tag{3.3}$$

This means that

$$\int_0^{+\infty}\!\!\int_{\mathbb{R}} \big(E(u)\,\varphi_t + F(u)\,\varphi_x\big)\,dt\,dx \,+\, \int_{\mathbb{R}} E(u^0(x))\,\varphi(0,x)\,dx \,\ge 0$$

for all nonnegative φ in the space $\mathcal{C}_c^\infty([0,+\infty[\times\mathbb{R})$ of $\mathcal{C}^\infty$ functions on $[0,+\infty[\times\mathbb{R}$ with compact support.

We shall also consider classical solutions, that is, functions $u = u(t,x)$ in $\mathcal{C}^1(]0,+\infty[\times\mathbb{R})\cap \mathcal{C}([0,+\infty[\times\mathbb{R})$ satisfying (3.1) pointwise.

In particular any classical solution to (3.1) satisfies (3.2), i.e. is an entropy solution, and conversely any entropy solution satisfies (3.1) in the distribution sense.

For entropy solutions, the following result is due to Kružkov (see [65]) :

Theorem 3.1. *For every $u^0 \in L^\infty(\mathbb{R})$, there exists a unique entropy solution u to* (3.1) *in $L^\infty(]0,+\infty[\times\mathbb{R}) \cap \mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$.*

Moreover for classical solutions, we have (see [101] for instance) :

Theorem 3.2. *Given a $\mathcal{C}^2$ flux f and a $\mathcal{C}^1$ bounded initial datum u^0 such that $f' \circ u^0$ be nondecreasing on $\mathbb{R}$, the unique entropy solution u to* (3.1) *is a classical solution.*

In this work we shall consider initial data in the subset $\mathcal{U}$ of $L^\infty(\mathbb{R})$ defined by

Definition 3.3. A function $v : \mathbb{R} \to \mathbb{R}$ belongs to $\mathcal{U}$ if it is nondecreasing, right-continuous, and has limits 0 and 1 at $-\infty$ and $+\infty$ respectively.

The following proposition expresses that this set is preserved by the conservation law (3.1) :

Proposition 3.4. *Given an initial datum $u^0 \in \mathcal{U}$, the entropy solution u given by Theorem 3.1 is such that $u(t, .)$ belongs to $\mathcal{U}$ for all $t \geq 0$.*

More precisely, given any $t \geq 0$, the $L^\infty(\mathbb{R})$ function $u(t, .)$ is *a.e.* equal to an element of the set $\mathcal{U}$.

Proof. On one hand u^0 has bounded variation, and thus so has $u(t, .)$ for all $t \geq 0$ by a general property of the entropy solution. In particular $u(t, .)$ is continuous but at countably many points, at which it has left and right limits. At each x we may define $u(t, x)$ by its right limit, which makes $u(t, .)$ right-continuous on the whole $\mathbb{R}$.

On the other hand, according to two general properties of the entropy solution, possible limits at $-\infty$ and $+\infty$ and monotonicity properties are preserved by the conservation law. In particular here $u(t, .)$ has limits 0 and 1 at $-\infty$ and $+\infty$ respectively, and is nondecreasing. This concludes the proof of Proposition 3.4. □

The set $\mathcal{U}$ is characterized by

Proposition 3.5. *The distributional derivative v_x of any $v \in \mathcal{U}$ is a Borel probability measure on $\mathbb{R}$, and for any $x \in \mathbb{R}$,*

$$v(x) \;=\; v_x(]-\infty, x]).$$

Conversely, if μ is a probability measure on $\mathbb{R}$, then v defined on $\mathbb{R}$ as

$$v(x) \;=\; \mu(]-\infty, x])$$

belongs to $\mathcal{U}$, and $v_x = \mu$.

Consequently the map $v \mapsto v_x$ is one-to-one from $\mathcal{U}$ onto the set $\mathcal{P}$ of probability measures on $\mathbb{R}$ (and $\mathcal{U}$ can be seen as the set of repartition functions of real-valued random variables).

Propositions 3.4 and 3.5 allow us to characterize at any time the distance between two solutions (with initial datum in $\mathcal{U}$) in terms of their space derivatives, in particular by means of the Wasserstein distances : given any real number $p \geq 1$, the Wasserstein distance of order p is defined on the set of probability measures on $\mathbb{R}$ by

$$W_p(\mu, \tilde{\mu}) = \inf_\pi \Big(\int_{\mathbb{R}^2} |x-y|^p \, d\pi(x,y) \Big)^{1/p}$$

where π runs over the set of probability measures on $\mathbb{R}^2$ with marginals μ and $\tilde{\mu}$; these distances are considered here in a broad sense with possibly infinite values.

This paper aims at proving that the Wasserstein distances between the space derivatives of any two such entropy solutions is a nonincreasing function of time :

Theorem 3.6. *Given a locally Lipschitz real-valued function f on $\mathbb{R}$ and two initial data u^0 and $\tilde{u}^0$ in $\mathcal{U}$, let u and $\tilde{u}$ be the associated entropy solutions to* (3.1)*. Then, for any $t \geq 0$ and $p \geq 1$, we have (with possibly infinite values)*

$$W_p(u_x(t,.), \tilde{u}_x(t,.)) \leq W_p(u^0_x, \tilde{u}^0_x).$$

We shall see in Section 3.2 that for $p = 1$ the distance W_1 satisfies

$$W_1(v_x, \tilde{v}_x) = \|v - \tilde{v}\|_{L^1(\mathbb{R})}$$

for all $v, \tilde{v} \in \mathcal{U}$. Hence Theorem 3.6 reads in the case $p = 1$:

$$\|u(t,.) - \tilde{u}(t,.)\|_{L^1(\mathbb{R})} \leq \|u^0 - \tilde{u}^0\|_{L^1(\mathbb{R})}.$$

Thus, for initial profiles in $\mathcal{U}$, we recover the L^1-contraction property given by Kružkov.

The result of Theorem 3.6 can be improved in the case of classical solutions, since in this case the Wasserstein distance between two solutions is conserved :

Theorem 3.7. *Given a $\mathcal{C}^1$ real-valued function f on $\mathbb{R}$, let u^0 and $\tilde{u}^0$ in $\mathcal{U}$ be two initial data such that the associated entropy solutions u and $\tilde{u}$ to* (3.1) *be classical solutions, increasing in x for all $t \geq 0$. Then for any $t \geq 0$ and $p \geq 1$ we have (with possibly infinite values)*

$$W_p(u_x(t,.), \tilde{u}_x(t,.)) = W_p(u^0_x, \tilde{u}^0_x).$$

From these general results can be induced some corollaries in the case of initial data in the subsets $\mathcal{U}_p$ of $\mathcal{U}$ defined as :

Definition 3.8. Let $p \geq 1$. A function v in $\mathcal{U}$ belongs to $\mathcal{U}_p$ if its distributional derivative v_x has finite moment of order p, that is, if $\displaystyle\int_{\mathbb{R}} |x|^p \, dv_x(x)$ is finite.

As in Proposition 3.5 the map $v \mapsto v_x$ is one-to-one from $\mathcal{U}_p$ onto the set $\mathcal{P}_p$ of probability measures on $\mathbb{R}$ with finite moment of order p. But we shall note in Section 3.2 that the map W_p on $\mathcal{P}_p \times \mathcal{P}_p$ defines a distance on $\mathcal{P}_p$. Then the real-valued map d_p defined on $\mathcal{U}_p \times \mathcal{U}_p$ by

$$d_p(v, \tilde{v}) = W_p(v_x, \tilde{v}_x)$$

induces a distance on $\mathcal{U}_p$, and for the associated topology we have

Corollary 3.9. *Given a locally Lipschitz function f on $\mathbb{R}$, $p \geq 1$ and $u^0 \in \mathcal{U}_p$, the entropy solution u to* (3.1) *belongs to $\mathcal{C}([0, +\infty[, \mathcal{U}_p)$.*

In particular for $p = 1$

$$d_1(v, \tilde{v}) = W_1(v_x, \tilde{v}_x) = \|v - \tilde{v}\|_{L^1(\mathbb{R})},$$

and the previous result can be precised by

Corollary 3.10. *Given a locally Lipschitz function f on $\mathbb{R}$ and $u^0 \in \mathcal{U}_1$, the entropy solution u to* (3.1) *is such that*

$$\|u(t,.) - u(s,.)\|_{L^1(\mathbb{R})} \leq |t - s| \, \|f'\|_{L^\infty(]0,1[)}.$$

This known result holds under weaker assumptions (for u^0 with bounded variation, see [101]), but in our case it will be recovered in a straightforward way.

Finally Theorem 3.7 can be precised in the $\mathcal{U}_p$ framework in the following way :

Corollary 3.11. *Given a $\mathcal{C}^2$ convex flux f and two $\mathcal{C}^1$ increasing initial data u^0 and $\tilde{u}^0$ in $\mathcal{U}_p$ for some $p \geq 1$, the following three properties hold :*

1. *the associated entropy solutions u and $\tilde{u}$ are classical solutions ;*
2. *$u(t,.)$ and $\tilde{u}(t,.)$ belong to $\mathcal{U}_p$ and are increasing for all $t \geq 0$;*
3. *for all $t \geq 0$, we have (with finite values)*

$$W_p(u_x(t,.), \tilde{u}_x(t,.)) \;=\; W_p(u_x^0, \tilde{u}_x^0).$$

The paper is organized as follows. The definition and some properties of Wasserstein distances are discussed in greater detail in Section 3.2. In Section 3.3 we consider the case of classical solutions, proving Theorem 3.7 and Corollary 3.11. Then the general case of entropy solutions is studied in Sections 3.4 and 3.5 : more precisely in Section 3.4 we introduce a time-discretized scheme, show the W_p contraction property for this discretized evolution and prove the convergence of the corresponding approximate solution toward the entropy solution ; Theorem 3.6 and its corollaries follow from this in Section 3.5. In Section 3.6 we shall finally see how the contraction property given in Theorem 3.6 extends to viscous conservation laws, whereas a generalization to solutions of conservation laws with fifferent flux functions is proposed in Section 3.7.

3.2 Wasserstein distances

In this section p is a real number with $p \geq 1$, $\mathcal{P}$ (resp. $\mathcal{P}_p$) stands for the set of probability measures on $\mathbb{R}$ (resp. with finite moment of order p) and dx for the Lebesgue measure on $\mathbb{R}$.

The Wasserstein distance of order p, valued in $\mathbb{R} \cup \{+\infty\}$, is defined on $\mathcal{P} \times \mathcal{P}$ by

$$W_p(\mu, \tilde{\mu}) \;=\; \inf_\pi \Big(\int_{\mathbb{R}^2} |x - y|^p \, d\pi(x,y) \Big)^{1/p} \tag{3.4}$$

where π runs over the set of probability measures on $\mathbb{R}^2$ with marginals μ and $\tilde{\mu}$. It is equivalently defined by

$$W_p(\mu,\tilde{\mu}) = \inf_{X_\mu, X_{\tilde{\mu}}} \Big(\int_0^1 |X_\mu(w) - X_{\tilde{\mu}}(w)|^p \, dw\Big)^{1/p} \tag{3.5}$$

where the infimum is taken over all random variables X_μ and $X_{\tilde{\mu}}$ on the probability space $(]0,1[, dw)$ with respective laws μ and $\tilde{\mu}$. It takes finite values on $\mathcal{P}_p \times \mathcal{P}_p$ and indeed defines a distance on $\mathcal{P}_p$.

For complete references about the Wasserstein distances and related topics the reader can refer to [111]. We only mention that both infima in (3.4) and (3.5) are achieved, and for the second definition we shall precise some random variables that achieve the infimum. For this purpose we introduce the notion of generalized inverse :

Definition 3.12. Let v belong to $\mathcal{U}$. Then its generalized inverse is the function v^{-1} defined on $]0,1[$ by

$$v^{-1}(w) \;=\; \inf\{x \in \mathbb{R};\, v(x) > w\}.$$

Then v^{-1} is a nondecreasing random variable on $(]0,1[, dw)$ by definition, with law v_x since

$$\begin{aligned}
\int_0^1 f(v^{-1}(w))\, dw &= \int_0^1 \Big(\int_{\mathbb{R}} f'(s) \mathbf{1}_{\{s \leq v^{-1}(w)\}}\, ds\Big)\, dw \\
&= \int_{\mathbb{R}} \Big(\int_0^1 f'(s)\mathbf{1}_{\{v(s) \leq w\}}\, dw\Big)\, ds \\
&= \int_{\mathbb{R}} f'(s)(1 - v(s))\, ds \\
&= \int_{\mathbb{R}} f(s)\, dv_x(s)
\end{aligned}$$

for all f in $\mathcal{C}^1_c(\mathbb{R})$. In particular its repartition function is v.

Moreover this generalized inverse achieves the infimum in (3.5) :

Proposition 3.13. *Let v and $\tilde{v}$ in $\mathcal{U}$. Then we have (with possibly infinite values)*

$$W_p(v_x, \tilde{v}_x) \;=\; \Big(\int_0^1 |v^{-1}(w) - \tilde{v}^{-1}(w)|^p\, dw\Big)^{1/p}$$

for all $p \geq 1$. In particular for $p = 1$ we also have

$$W_1(v_x, \tilde{v}_x) \;=\; \|v - \tilde{v}\|_{L^1(\mathbb{R})}.$$

Proof. The general result is proved in [111]. The result specific to the case $p = 1$ follows by introducing, for a given $v \in \mathcal{U}$, the map defined on $\mathbb{R}\times]0,1[$ by

$$jv(x,w) = \begin{cases} 1 & \text{if } v(x) > w \\ 0 & \text{if } v(x) \leq w, \end{cases}$$

for which we have

$$|v^{-1} - \tilde{v}^{-1}|(w) = \int_{\mathbb{R}} |jv - j\tilde{v}|(x,w)\ dx$$

for almost every $w \in]0,1[$, and

$$\int_0^1 |jv - j\tilde{v}|(x,w)\ dw \ = \ |v - \tilde{v}|(x)$$

for almost every $x \in \mathbb{R}$. Integrating the first equality on w in $]0,1[$ and the second one on x in $\mathbb{R}$, we deduce

$$\int_0^1 |v^{-1} - \tilde{v}^{-1}|(w)\ dw = \int_{\mathbb{R}} |v - \tilde{v}|(x)\ dx.$$

□

Given $v \in \mathcal{U}$, its generalized inverse v^{-1} is actually the a.e. unique nondecreasing random variable on $(]0,1[, dw)$ with law v_x. Given any other random variable X on $(]0,1[, dw)$ with law v_x, v^{-1} is called the (a.e. unique) nondecreasing rearrangement of X (see [111]).

We conclude this section recalling a result relative to the convergence of probability measures. A sequence (μ_n) of probability measures on $\mathbb{R}$ is said to converge weakly toward a probability measure μ if, as n goes to $+\infty$, $\int_{\mathbb{R}} \varphi\, d\mu_n$ tends to $\int_{\mathbb{R}} \varphi\, d\mu$ for all bounded continuous real-valued functions φ on $\mathbb{R}$ (or equivalently for all $\mathcal{C}^\infty$ functions φ with compact support, that is, if μ_n converges to μ in the distribution sense). Given $p \geq 1$ this convergence is metrized on $\mathcal{P}_p$ by the distance W_p as shown by the following proposition (see [111]) :

Proposition 3.14. *Let $p \geq 1$, (μ_n) a sequence of probability measures in $\mathcal{P}_p$ and $\mu \in \mathcal{P}$. Then the following statements are equivalent :*

i) $(W_p(\mu_n, \mu))$ converges to 0 ;

ii) (μ_n) converges weakly to μ and $\sup_n \int_{|x| \geq R} |x|^p\, d\mu_n(x)$ tends to 0 as R goes to infinity.

In this proposition we do not a priori assume that μ belongs to $\mathcal{P}_p$, but it can be noted that this property is actually induced by any of both hypotheses *i*) and *ii*).

For measures in $\mathcal{P}$ we have the weaker result :

Proposition 3.15. *Let $p \geq 1$, (μ_n) and (ν_n) two sequences in $\mathcal{P}$ converging weakly to μ and ν in $\mathcal{P}$ respectively. Then (with possibly infinite values)*

$$W_p(\mu, \nu) \leq \liminf_{n \to +\infty} W_p(\mu_n, \nu_n).$$

3.3 The case of classical solutions : Theorem 3.7 and corollary

3.3.1 Proof of Theorem 3.7

We consider two classical solutions u and $\tilde{u}$ to (3.1) such that $u(t,.)$ and $\tilde{u}(t,.)$ belong to $\mathcal{U}$ and be increasing for all $t \geq 0$, and we shall prove that

$$W_p(u_x(t,.), \tilde{u}_x(t,.)) \, = \, W_p(u_x^0, \tilde{u}_x^0)$$

as a consequence of Proposition 3.13.

The map u^0 is increasing from 0 to 1, so has a (true) inverse $X(0,.)$ defined on $]0,1[$ by

$$u^0(X(0,w)) \, = \, w.$$

Then, given $w \in]0,1[$, we consider a characteristic curve $t \mapsto X(t,w)$ solution of

$$X_t(t,w) = f'\big(u(t, X(t,w))\big) \tag{3.6}$$

for $t \geq 0$, and taking value $X(0,w)$ at $t = 0$. Since f is $\mathcal{C}^1$ and u is bounded there exists a (non necessarily unique) solution $X(.,w)$ to (3.6) by Peano Theorem (see [61] for instance) ; moreover by a classical computation from (3.1) it is known to satisfy

$$u(t, X(t,w)) \, = \, w \tag{3.7}$$

for all $t \geq 0$, from which it follows that

$$X_t(t,w) \Big(= f'(u(t, X(t,w)))\Big) = f'(w)$$

and hence

$$X(t,w) \, = \, X(0,w) \, + \, t f'(w). \tag{3.8}$$

In particular there exists a unique solution $X(.,w)$ to (3.6). Now given $t \geq 0$, $X(t,.)$ is the (true) inverse of the increasing function $u(t,.)$ (by (3.7)), and Proposition 3.13 writes

$$W_p(u_x(t,.), \tilde{u}_x(t,.)) \, = \, \Big(\int_0^1 |X(t,w) - \tilde{X}(t,w)|^p \, dw \Big)^{1/p}.$$

But from (3.8) we obtain

$$X(t,w) \, - \, \tilde{X}(t,w) \, = \, X(0,w) \, - \, \tilde{X}(0,w). \tag{3.9}$$

This result ensures in particular that $W_p(u_x(t,.), \tilde{u}_x(t,.))$ remains constant in time, may its initial value be finite or not ; note however that (3.9) is actually much stronger that Theorem 3.7. $\square$

3.3.2 Proof of Corollary 3.11

We assume that f is a $\mathcal{C}^2$ convex function on $\mathbb{R}$, and u^0 is a $\mathcal{C}^1$ increasing initial profile in $\mathcal{U}_p$.

First of all we note that the associated entropy solutions u is a classical solution in view of Theorem 3.2 : this result is proved in [101] for instance, and its proof also ensures that $u(t,.)$ is increasing for all $t \geq 0$.

Then we check that the moment property is preserved by the conservation law, that is, that $u(t,.)$ also belongs to $\mathcal{U}_p$ for any $t \geq 0$. Indeed, given $t \geq 0$, we have by the change of variable $w = [u(t,.)](x)$:

$$\begin{aligned}
\int_{\mathbb{R}} |x|^p \, u_x(t,x)\, dx &= \int_0^1 |X(t,w)|^p \, dw \\
&= \int_0^1 |X(0,w) + t f'(w)|^p \, dw \\
&\leq 2^{p-1} \left[\int_0^1 |X(0,w)|^p \, dw \, + t^p \|f'\|^p_{L^\infty(]0,1[)} \right]
\end{aligned}$$

which is finite since

$$\int_0^1 |X(0,w)|^p \, dw = \int_{\mathbb{R}} |x|^p \, u^0_x(x)\, dx$$

is finite by assumption. This ends the proof of Corollary 3.11. □

3.4 Time discretization of the conservation law

In the previous section we have seen that the classical solutions are obtained through the method of characteristics, that we now summarize in our case : given an initial profile u^0 in $\mathcal{U}$ such that the corresponding solution u be $\mathcal{C}^1$ and increasing in x for all $t \geq 0$, let $X(0,.)$ be its inverse, defined by

$$u^0(X(0,w)) \, = \, w$$

for all $w \in]0,1[$. Let then $X(0,w)$ evolve into

$$X(t,w) \, = \, X(0,w) + t f'(w) \tag{3.10}$$

for all $t \geq 0$ and $w \in]0,1[$ (see (3.8)). The solution $u(t,.)$ is then the inverse of the increasing map $X(t,.)$, that is, is the unique solution of

$$X(t, u(t,x)) \, = \, x.$$

In the general case, defining $X(0,.)$ in some similar way, there is no hope for the function $X(t,.)$ defined by (3.10) to be increasing for $t > 0$; inverting it would thus lead to a multivalued function, and no more to the entropy solution of the conservation law, as in the particular case discussed above.

However, averaging (or "collapsing") this multivalued function into a single-valued function, Y. Brenier showed in [26] how to build an approximate solution to the conservation law.

We now precisely describe this so-called Transport-Collapse method in our case.

3.4.1 Definition and W_p contraction property of the discretized solution

Let $u^0 \in \mathcal{U}$ be some fixed initial profile, with generalized inverse $X(0,.)$ given as in Definition 3.12 by

$$X(0,w) \;=\; \inf\{x \in \mathbb{R}; u^0(x) > w\}$$

for all $w \in]0,1[$. $X(0,.)$ can be seen as a random variable on the probability space $]0,1[$ equipped with the Lebesgue measure dw ; its law is u^0_x, as pointed out after Definition 3.12.

We let then $X(0,.)$ evolve according to the method of characteristics, denoting

$$X(h,w) \;=\; X(0,w) \,+\, hf'(w)$$

for all $h \geq 0$ and almost every $w \in]0,1[$. Again, given $h \geq 0$, $X(h,.)$ can be seen as a random variable on $]0,1[$; let then $T_h u^0$ be its repartition function, that is, the function belonging to $\mathcal{U}$ and defined at any $x \in \mathbb{R}$ as the Lebesgue measure of the set $\{w \in]0,1[; X(h,w) \leq x\}$. It is given by

$$T_h u^0(x) \;=\; \int_0^1 \mathbf{1}_{\{X(h,w)\leq x\}}(w)\, dw.$$

We summarize this construction in the following definition :

Definition 3.16. Let $v \in \mathcal{U}$ with generalized inverse $X(0,.)$ defined on $]0,1[$ by

$$X(0,w) \;=\; \inf\{x \in \mathbb{R}; v(x) > w\}.$$

Then, given $h > 0$, and letting

$$X(h,w) \;=\; X(0,w) \,+\, hf'(w)$$

for almost every $w \in]0,1[$, we define the $\mathcal{U}$ function $T_h v$ on $\mathbb{R}$ by

$$T_h v(x) \;=\; \int_0^1 \mathbf{1}_{\{X(h,w)\leq x\}}(w)\, dw.$$

In the case of Section 3.3 (see (3.8)), it turns out that $X(h,.)$ is the (true) inverse of $T_h u^0$, and $(h,x) \mapsto T_h u^0(x)$ is exactly the entropy solution to equation (3.1) with initial datum u^0 in $\mathcal{U}$. This does not hold anymore in the general case, but will allow us to build an approximate solution $S_h u^0$ by iterating the operator T_h. Let us first give two important properties of T_h :

Proposition 3.17. *Let $h > 0$, T_h defined as above and $p \geq 1$. Then*
i) $T_h v$ belongs to $\mathcal{U}_p$ if so does v.
ii) For any v and $\tilde{v}$ in $\mathcal{U}$ we have (with possibly infinite values unless v and $\tilde{v} \in \mathcal{U}_p$)

$$W_p([T_h v]_x, [T_h \tilde{v}]_x) \;\leq\; W_p(v_x, \tilde{v}_x).$$

Proof. It is really similar to what has been done in Section 3.3 as for Corollary 3.11.
i) $T_h v$ belongs to $\mathcal{U}$ as a repartition function of a random variable, and we have

$$\begin{aligned}
\int_{\mathbb{R}} |x|^p \, d[T_h v]_x(x) &= \int_0^1 |X(h,w)|^p \, dw \\
&= \int_0^1 |X(0,w) \,+\, hf'(w)|^p \, dw \\
&\leq 2^{p-1} \int_0^1 |X(0,w)|^p + |hf'(w)|^p \, dw \\
&\leq 2^{p-1} \left[\int_{\mathbb{R}} |x|^p \, dv_x(x) + h^p \|f'\|^p_{L^\infty(]0,1[)} \right],
\end{aligned}$$

which ensures that $[T_h v]_x$ has finite moment of order p if so does v_x.

ii) On one hand the generalized inverses $X(0,.)$ and $\tilde{X}(0,.)$ of v and $\tilde{v}$ respectively satisfy

$$W_p(v_x, \tilde{v}_x) \;=\; \Big(\int_0^1 |X(0,w) - \tilde{X}(0,w)|^p \, dw \Big)^{1/p} \tag{3.11}$$

by Proposition 3.13 (with finite values if both v and $\tilde{v}$ belong to $\mathcal{U}_p$, and possibly infinite otherwise). On the other hand $X(h,.)$ and $\tilde{X}(h,.)$ have respective law $[T_h v]_x$ and $[T_h \tilde{v}]_x$, so

$$W_p([T_h u^0]_x, [T_h \tilde{u}^0]_x) \;\leq\; \Big(\int_0^1 |X(h,w) - \tilde{X}(h,w)|^p \, dw \Big)^{1/p} \tag{3.12}$$

by definition of the Wasserstein distance. But

$$X(h,w) - \,\tilde{X}(h,w) \;=\; X(0,w) \;-\; \tilde{X}(0,w)$$

for almost every $w \in \,]0,1[$ by definition, which concludes the argument by (3.11) and (3.12).
Note again that (3.12) holds only as an inequality since $X(h,.)$ and $\tilde{X}(h,.)$ are not necessarily nondecreasing, which was the case in the example discussed in Section 3.3. □

We now use the operator T_h defined above to build an approximate solution $S_h u^0$ to the conservation law (3.1) :

Definition 3.18. Let h be some positive number and $v \in \mathcal{U}$. For any $t \geq 0$ decomposed as $t = (N+n)h$ with $N \in \mathbb{N}$ and $0 \leq n < 1$, we let

$$S_h v(t,.) \;=\; (1-n)\, T_h^N v(.) \;+\; n\, T_h^{N+1} v(.)$$

where $T_h^0 v = v$ and $T_h^{N+1} v = T_h(T_h^N v)$.

These iterations make sense because $T_h v \in \mathcal{U}$ if $v \in \mathcal{U}$, $S_h v(t,.) \in \mathcal{U}$ (resp. $\mathcal{U}_p$) for any $h, t \geq 0$ and $v \in \mathcal{U}$ (resp. $\mathcal{U}_p$).

We now prove two contractions properties on these approximate solutions. We first have the W_p contraction property :

Proposition 3.19. *Let h be some fixed positive number and S_h defined as above. Then, given v and $\tilde{v}$ in $\mathcal{U}$, we have for any $t \geq 0$:*

$$W_p([S_h v]_x(t,.), [S_h \tilde{v}]_x(t,.)) \leq W_p(v_x, \tilde{v}_x).$$

Proof. It follows from Proposition 3.17 (about T_h) and to the convexity of the W_p distance to the power p, in the sense that

$$W_p^p(\alpha\mu_1 + (1-\alpha)\mu_2, \alpha\nu_1 + (1-\alpha)\nu_2) \leq \alpha W_p^p(\mu_1, \nu_1) + (1-\alpha) W_p^p(\mu_2, \nu_2)$$

for all real number $\alpha \in [0,1]$ and probability measures μ_1, μ_2, ν_1 and ν_2 (see [111] for instance). $\square$

Then we have the $L^1(\mathbb{R})$ contraction property :

Proposition 3.20. *Let h be some fixed positive number and S_h defined as above. Then, for any $v \in \mathcal{U}$ and $s, t \geq 0$ we have*

$$\|S_h v(t,.) - S_h v(s,.)\|_{L^1(\mathbb{R})} \leq |t-s| \, \|f'\|_{L^\infty(]0,1[)}.$$

Proof. We first observe that

$$\|T_h V - V\|_{L^1(\mathbb{R})} \leq h \|f'\|_{L^\infty(]0,1[)} \tag{3.13}$$

for any $V \in \mathcal{U}$. Indeed

$$T_h V(x) = \int_0^1 \mathbf{1}_{\{V^{-1}(w) + hf'(w) \leq x\}} \, dw$$

for all $x \in \mathbb{R}$ by definition, so for any test function $\phi \in \mathcal{C}_c(\mathbb{R})$ we have

$$\begin{aligned} \int_{\mathbb{R}} [T_h V(x) - V(x)] \, \phi(x) \, dx &= \int_{\mathbb{R}} \Big(\int_0^1 [\mathbf{1}_{\{V^{-1}(w)+hf'(w) \leq x\}} - \mathbf{1}_{\{V^{-1}(w) \leq x\}}] \, dw \Big) \phi(x) \, dx \\ &= \int_0^1 \Big(\int_{\mathbb{R}} [\mathbf{1}_{\{V^{-1}(w)+hf'(w) \leq x\}} - \mathbf{1}_{\{V^{-1}(w) \leq x\}}] \, \phi(x) \, dx \Big) \, dw \\ &\leq h \, \|f'\|_{L^\infty(]0,1[)} \, \|\phi\|_{L^\infty(\mathbb{R})}, \end{aligned}$$

which yields the announced bound on $\|T_h V - V\|_{L^1(\mathbb{R})}$.

Then we let $s = (M+m)h$ and $t = (N+n)h$ with $M, N \in \mathbb{N}$ and $0 \leq m, n < 1$, and, for example assuming that $N > M$, considering the intermediate times $(M+1)h, \ldots, Nh$, we have

$$S_h v(t,.) - S_h v(s,.) = n \, [T_h(T_h^N v) - T_h^N v] + \sum_{k=M+1}^{N-1} [T_h(T_h^k v) - T_h^k v] + (1-m)[T_h(T_h^M v) - T_h^M v].$$

Then, as $T_h^k v$ belongs to $\mathcal{U}$ for all integer k, we have from the preliminary estimate :

$$\|S_h v(t,.) - S_h v(s,.)\|_{L^1(\mathbb{R})} \le [(1-m) + \sum_{k=M+1}^{N-1} 1 \ + n]\, h\, \|f'\|_{L^\infty(]0,1[)} = (t-s)\|f'\|_{L^\infty(]0,1[)}.$$

□

We shall now make use of this $L^1(\mathbb{R})$ contraction property to prove the convergence of the scheme toward the entropy solution of the conservation law.

3.4.2 Convergence of the scheme in the $L^1_{loc}(\mathbb{R})$ sense

In this section we prove

Proposition 3.21. *Let $u^0 \in \mathcal{U}$. Then, as h goes to 0, the function $S_h u^0$ converges in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$ to the entropy solution of (3.1) with initial datum u^0.*

Our proof, which will go in several steps, follows the one of Brenier in [26], adapted to functions of $\mathcal{U}$ instead of $L^1(\mathbb{R})$.

1. Compactness of the sequence

We give here a compactness result on the family of approximate solutions. For this purpose we first recall that $L^1_{loc}(\mathbb{R})$ is equipped with the topology defined by the semi-norms

$$p_m(f) \;=\; \int_{-m}^{m} |f(x)|\, dx$$

for any integer m and $f \in L^1_{loc}(\mathbb{R})$, and that it is metrizable for this topology.

Then the space $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$ is equipped with the topology defined by the semi-norms

$$q_{nm}(f) \;=\; \sup_{t\in[0,n]} p_m(f)$$

for any integers n and m and $f \in \mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$, and this topology is induced by a metric.

In particular a family $\mathcal{F}$ is relatively compact in $L^1_{loc}(\mathbb{R})$ $\big($resp. $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))\big)$ if and only if any sequence in $\mathcal{F}$ has a subsequence converging in $L^1_{loc}(\mathbb{R})$ $\big($resp. $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))\big)$.

Then we shall prove

Proposition 3.22. *Given $v \in \mathcal{U}$ the family $(S_h v)_h$ is relatively compact in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$.*

Proof. On one hand the family $(t \mapsto S_h v(t,.))_h$ is uniformly equicontinuous from $[0,+\infty[$ into $L^1_{loc}(\mathbb{R})$ by Proposition 3.20.

On the other hand, given $t \ge 0$, the family $(S_h v(t,.))_h$ is relatively compact in $L^1_{loc}(\mathbb{R})$. Indeed, as previously noted, any function in $\mathcal{U}$ is bounded by 1 and belongs to the set $BV(\mathbb{R})$ of real-valued $L^1_{loc}(\mathbb{R})$ functions with bounded variation, with total variation equal to 1. But

$S_h v(t,.) \in \mathcal{U}$ for any h, so the family $(S_h v(t,.))_h$ is bounded in $L^\infty(\mathbb{R}) \cap BV(\mathbb{R})$, and thus as announced is relatively compact in $L^1_{loc}(\mathbb{R})$ by Helly's theorem.

Arzela-Ascoli Theorem (see [99] for instance) then ensures that the family $(S_h v)_h$ is relatively compact in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$. □

2. The entropy inequality : proof of Proposition 3.21

In this section we prove that, given $u^0 \in \mathcal{U}$, the limit of any sequence of $(S_h u^0)_h$ converging in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$ is the entropy solution to the conservation law (3.1) with initial datum u^0. By the compactness property of the family $(S_h u^0)_h$ established in Proposition 3.22 this shall ensure Proposition 3.21

We have noted that if the flux f and the initial profile $u^0 \in \mathcal{U}$ are such that the entropy solution u of the conservation law (3.1) is a classical increasing (in x) solution, then actually

$$u(t,.) = T_t u^0$$

for all $t \geq 0$. In particular

$$[E(T_t u^0)]_t(x) + [F(T_t u^0)]_x(x) \leq 0$$

for any Lipschitz convex function E and associated flux F defined by (3.3). This formally induces

$$\frac{d}{dt} \int_{\mathbb{R}} E(T_t u^0)(x)\, \phi(x)\, dx - \int_{\mathbb{R}} F(T_t u^0)(x)\, \phi'(x)\, dx \leq 0$$

for any nonnegative test function $\phi \in \mathcal{C}^1_c(\mathbb{R})$, and in particular at $t = 0$

$$\frac{d}{dt} \int_{\mathbb{R}} E(T_t u^0)(x)\, \phi(x)\, dx \Big|_{t=0} \leq \int_{\mathbb{R}} F(u^0)(x)\, \phi'(x)\, dx. \tag{3.14}$$

In the general case where the initial profile u^0 is any function in $\mathcal{U}$, the following discrete version of (3.14) holds :

Proposition 3.23. *Let E be a Lipschitz convex function, with associated flux F defined as in* (3.3)*, $v \in \mathcal{U}, h \geq 0$ and $\phi \in \mathcal{C}^2_c(\mathbb{R})$, nonnegative and with support included in $[-R,+R]$. Then*

$$\int_{\mathbb{R}} [E(T_h v(x)) - E(v(x))]\, \phi(x)\, dx$$
$$\leq h \int_{\mathbb{R}} F(v(x))\, \phi'(x)\, dx + h^2 \|\phi''\|_{L^\infty} \|f'\|^2_{L^\infty(]0,1[)} \|E'\|_{L^\infty} (R + h\|f'\|_{L^\infty(]0,1[)}).$$

Assuming this result for the moment we show how it implies Proposition 3.21.

Given $u^0 \in \mathcal{U}$, we consider a sequence of the family $(S_h u^0)_h$ converging to a function u in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$, and we shall prove from Proposition 3.23 that u is a solution to the conservation law with initial datum u^0, that is, that $u \in L^\infty(]0,+\infty[\times\mathbb{R})$ and satisfies

$$\int_0^{+\infty}\int_{\mathbb{R}}\Big(E(u(t,x))\,\varphi_t(t,x)+F(u(t,x))\,\varphi_x(t,x)\Big)dt\,dx\;+\;\int_{\mathbb{R}}E(u^0(x))\,\varphi(0,x)\,dx\;\geq 0$$

for all nonnegative $\varphi\in\mathcal{C}_c^\infty([0,+\infty[\times\mathbb{R})$.

For this purpose we again denote $(S_hu^0)_h$ the converging sequence in $\mathcal{C}([0,+\infty[,L^1_{loc}(\mathbb{R}))$. This sequence is bounded in $L^\infty(]0,+\infty[\times\mathbb{R})$, and thus its limit u belongs to $L^\infty(]0,+\infty[\times\mathbb{R})$.

Let now $\varphi\in\mathcal{C}_c^\infty([0,+\infty[\times\mathbb{R})$, be nonnegative and such that $\varphi(t,x)\equiv 0$ for $t\geq T$ or $|x|\geq R$. Given $h\geq 0$ we denote $N=[T/h]$ and apply Proposition 3.23 to $v=S_hu^0(kh,.)$ and $\phi=\varphi(kh,.)$ for $k=0,\ldots,N$. Letting $M=\|f'\|_{L^\infty(]0,1[)}$ and summing on $k=0,\ldots,N$ yield then by a change of indexes :

$$\begin{aligned}\sum_{k=1}^{N}\int_{\mathbb{R}}E(S_hu^0(kh,x))\;[\varphi((k-1)h,x)-\varphi(kh,x)]\;dx&\\ +\int_{\mathbb{R}}E(S_hu^0((N+1)h,x))\,\varphi(Nh,x)\,dx\;-\int_{\mathbb{R}}E(u^0(x))\varphi(0,x)\,dx&\\ \leq h\sum_{k=0}^{N}\int_{\mathbb{R}}F(S_hu^0(kh,x))\,\varphi_x(kh,x)\,dx+h^2\sum_{k=0}^{N}\|\varphi_{xx}(kh,.)\|_{L^\infty(\mathbb{R})}(R+hM)M^2\|E'\|_{L^\infty}.&\end{aligned}$$

First of all the second terms in both sides converge to 0 as h goes to 0. Then

$$\begin{aligned}\sum_{k=1}^{N}\int_{\mathbb{R}}E(S_hu^0(kh,x))\;\;[\varphi(kh,x)-\varphi((k-1)h,x)]\,dx\;-\int_0^T\int_{\mathbb{R}}E(u(t,x))\,\varphi_t(t,x)\,dt\,dx&\\ =\sum_{k=1}^{N}\int_{\mathbb{R}}\int_{(k-1)h}^{kh}[E(S_hu^0(kh,x))-E(S_hu^0(t,x))]\varphi_t(t,x)\,dt\,dx&\\ +\sum_{k=1}^{N}\int_{\mathbb{R}}\int_{(k-1)h}^{kh}[E(S_hu^0(t,x))-E(u(t,x))]\varphi_t(t,x)\,dtdx+\int_{Nh}^{T}\int_{\mathbb{R}}E(u(t,x))\,\varphi_t(t,x)\,dt\,dx.&\end{aligned}$$

But by Proposition 3.20 the first term in the right-hand side is bounded by

$$\|E'\|_{L^\infty}M\|\varphi_t\|_{L^\infty}\sum_{k=1}^{N}\int_{(k-1)h}^{kh}|kh-t|\;dt\leq\|E'\|_{L^\infty}M\|\varphi_t\|_{L^\infty}\,Th,$$

the second term by

$$\begin{aligned}\|E'\|_{L^\infty}\|\varphi_t\|_{L^\infty}\sum_{k=1}^{N}\int_{(k-1)h}^{kh}\|S_hu^0(t,x)-u(t,x)\|_{L^1(]-R,R[)}dt&\\ \leq\|E'\|_{L^\infty}\|\varphi_t\|_{L^\infty}\,T\,\|S_hu^0(t,x)-u(t,x)\|_{\mathcal{C}([0,T],L^1(]-R,R[))},&\end{aligned}$$

and the third one by
$$2R\,\|E\|_{L^\infty([0,1]}\|\varphi_t\|_{L^\infty}\,h.$$
Hence the three of them converge to 0 as h goes to 0, which means that

$$\sum_{k=1}^{N}\int_{\mathbb{R}} E(S_h u^0(kh,x))\ \ [\varphi(kh,x)-\varphi((k-1)h,x)]\,dx\ \rightarrow\ \int_0^T\int_{\mathbb{R}} E(u(t,x))\,\varphi_t(t,x)\,dt\,dx.$$

We finally prove in the same way that

$$h\sum_{k=0}^{N}\int_{\mathbb{R}} F(S_h u^0(kh,x))\,\varphi_x(kh,x)\,dx\ \ \rightarrow\ \int_0^T\int_{\mathbb{R}} F(u(t,x))\,\varphi_(t,x)\,dt\,dx$$

as h goes to 0 by noting that F is bounded and Lipschitz on $[0,1]$.

Thus, at the limit, the entropy inequality (3.2) holds for u.

Consequently, given $u^0\in\mathcal{U}$, any sequence of $(S_h u^0)_{h\geq 0}$ has a subsequence converging in $\mathcal{C}([0,+\infty[,L^1_{loc}(\mathbb{R}))$ to a function which is an entropy solution to (3.1) with initial datum u^0. By the uniqueness of this solution ensured by Theorem 3.1 this concludes the proof of Proposition 3.21.

3. Proof of Proposition 3.23

We first give the following general lemma :

Lemma 3.24. *Let E be a Lipschitz convex function on $[0,1]$ and A a measurable subset of $([0,1],dw)$. Then*

$$E\Big(\int_0^1 \mathbf{1}_A(w)\,dw\Big)\ \leq\ E(0)\ +\ \int_0^1 \mathbf{1}_A(w)E'(w)\,dw.$$

Proof. We first assume that $E'(|A|)\ =\ 0$, where $|A|$ stands for the Lebesgue measure $\displaystyle\int_0^1 \mathbf{1}_A(w)\,dw$ of A.

Then E' is non-positive on $[0,|A|]$ and nonnegative on $[0,|A|]^c$. Thus on one hand

$$E(|A|)-E(0)=\int_{[0,|A|]}E'=\int_{[0,|A|]\cap A}E'+\int_{[0,|A|]\cap A^c}E'\leq\int_{[0,|A|]\cap A}E',$$

and on the other hand

$$\int_A E'=\int_{[0,|A|]\cap A}E'+\int_{[0,|A|]^c\cap A}E'\geq\int_{[0,|A|]\cap A}E',$$

which yields
$$E(|A|)-E(0)\leq\int_A E'(w)\,dw.$$

In the general case we replace E by $E - wE'(|A|)$ and apply the previous case. $\square$

We now turn to the **proof of Proposition** 3.23. For this, given $v \in \mathcal{U}$, we consider again the map jv defined on $\mathbb{R}\times]0,1[$ by

$$jv(x,w) = \begin{cases} 1 & \text{if } v(x) > w \\ 0 & \text{if } v(x) \leq w. \end{cases}$$

Lemma 3.24, applied to $A = \{w \in]0,1[; v(x - hf'(w)) > w\}$, yields

$$E\Big(\int_0^1 \mathbf{1}_{\{v(x-hf'(w))>w\}}\,dw\Big) \leq E(0) \,+\, \int_0^1 \mathbf{1}_{\{v(x-hf'(w))>w\}}E'(w)\,dw.$$

But

$$E(T_h v(x)) \,=\, E\Big(\int_0^1 \mathbf{1}_{\{v(x-hf'(w))>w\}}\,dw\Big)$$

by definition of $T_h v$, and

$$\int_0^1 \mathbf{1}_{\{v(x-hf'(w))>w\}}E'(w)\,dw \,=\, \int_0^1 jv(x-hf'(w),w)\,E'(w)\,dw,$$

$$E(v(x)) \,=\, E(0) + \int_0^1 jv(x,w)\,E'(w)\,dw,$$

by definition of jv, so

$$E(T_h v(x)) \,-\, E(v(x)) \,\leq\, \int_0^1 [jv(x-hf'(w),w) - jv(x,w)]\;E'(w)\,dw.$$

Consequently, for any nonnegative $\phi \in \mathcal{C}_c^2(\mathbb{R})$,

$$\int_{\mathbb{R}} [E(T_h v(x)) \,-\, E(v(x))]\;\phi(x)\,dx \leq \int_0^1\int_{\mathbb{R}} [\phi(x+hf'(w)) - \phi(x)]\,jv(x,w)E'(w)\,dw\,dx.$$

Then, assuming that f is M-Lipschitz on $[0,1]$, a Taylor expansion of ϕ around x yields

$$\phi(x+hf'(w)) - \phi(x) = hf'(w)\phi'(x) + \frac{1}{2}(hf'(w))^2\phi''(y(x,h,w))$$

for some $y(x,h,w) \in [x - hM, x + hM]$, and thus

$$\begin{aligned}\int_{\mathbb{R}} [E(T_h v(x)) \,-\, E(v(x))]\;\phi(x)\,dx \leq h\int_{\mathbb{R}}\Big(\int_0^1 jv(x,w)f'(w)E'(w)\,dw\Big)\,\phi'(x)\,dx \\ + \frac{h^2}{2}\int_0^1 f'(w)^2E'(w)\Big(\int_{\mathbb{R}}\phi''(y(x,h,w))jv(x,w)\,dx\Big)dw.\end{aligned}$$

But on one hand for the first term

$$\int_0^1 jv(x,w)f'(w)E'(w)\,dw = \int_0^1 jv(x,w)F'(w)\,dw = F(v(x)) - F(0) = F(v(x)).$$

On the other hand, for the second term, if ϕ has support in $[-R,R]$, then $\phi''(y) = 0$ for $y \in [x-hM, x+hM]$ and x such that $|x| \geq R+hM$, and hence

$$\int_0^1 f'(w)^2\, E'(w)\Big(\int_{\mathbb{R}} \phi''(y(x,h,w)jv(x,w)\,dx\Big)dw \leq M^2 \|E'\|_{L^\infty} \|\phi''\|_{L^\infty}\, 2(R+hM).$$

This concludes the proof of Proposition 3.23.

3.4.3 Convergence of the scheme in W_p distance sense

We first prove a uniform equiintegrability result on the approximate solutions :

Proposition 3.25. *Let S_h be defined as above, $v \in \mathcal{U}_p$ and $T \geq 0$. Then*

$$\sup_{0\leq h\leq T}\ \sup_{0\leq t\leq T} \int_{|x|\geq R} |x|^p d[S_h v]_x(t,x)$$

tends to 0 *as R goes to infinity.*

Proof. We again denote $M = \|f'\|_{L^\infty(]0,1[)}$, and first consider T_h itself, writing

$$\begin{aligned}
\int_{|x|\geq R} |x|^p d[T_h v]_x(x) &= \int_0^1 |v^{-1}(w) + hf'(w)|^p\, \mathbf{1}_{\{|v^{-1}(w)+hf'(w)|\geq R\}}\, dw \\
&\leq \int_{\mathbb{R}} (|x| + hM)^p \mathbf{1}_{\{|x|+hM\geq R\}}\, dv_x(x) \\
&\leq \left(1 + \frac{hM}{R-hM}\right)^p \int_{|x|\geq R-hM} |x|^p\, dv_x(x)
\end{aligned}$$

for $R > hM$. From this computation we deduce by iteration

$$\int_{|x|\geq R} |x|^p d[T_h^N v]_x(x) \leq \prod_{j=1}^N \left(1 + \frac{hM}{R-jhM}\right)^p \int_{|x|\geq R-NhM} |x|^p\, dv_x(x)$$

for $R > NhM$, with

$$\prod_{j=1}^N \left(1 + \frac{hM}{R-jhM}\right) \leq \left(1 + \frac{hM}{R-NhM}\right)^N \leq \exp\left(\frac{NhM}{R-NhM}\right).$$

Thus

$$\int_{|x|\geq R} |x|^p d[S_h v]_x(Nh,x) \leq \exp\left(\frac{pTM}{R-TM}\right) \int_{|x|\geq R-TM} |x|^p\, dv_x(x)$$

for any N and h such that $Nh \leq T$.

From this we get for instance

$$\int_{|x|\geq R} |x|^p d[S_h v]_x(t,x) \leq \exp\left(\frac{2pTM}{R-2TM}\right) \int_{|x|\geq R-2TM} |x|^p \, dv_x(x)$$

for any t and h smaller than T. This concludes the argument since the last integral tends to 0 as R goes to infinity. □

From this we deduce the convergence of the scheme in W_p distance sense :

Proposition 3.26. *Let $u^0 \in \mathcal{U}_p$ and u be the entropy solution to (3.1) with initial datum u^0. Then, for any $t \geq 0$, $W_p([S_h u^0]_x(t,.), u_x(t,.))$ converges to 0 as h goes to 0.*

Proof. Given $t \geq 0$, $S_h u^0(t,.)$ converges to $u(t,.)$ in $L^1_{loc}(\mathbb{R})$ as h goes to 0 (by Proposition 3.21), so $[S_h u]_x(t,.)$ converges to the probability measure $u_x(t,.)$, first in the distribution sense, then in the weak sense of probability measures, and finally in W_p distance by Propositions 3.25 and 3.14.

Note in particular that $u_x(t,.)$ has finite moment of order p for any $t \geq 0$, that is, $u(t,.)$ belongs to $\mathcal{U}_p$. □

3.5 The general case of entropy solutions : Theorem 3.6 and corollaries

3.5.1 Proof of Theorem 3.6

We let $p \geq 1$ and consider two initial data u^0 and $\tilde{u}^0$ in $\mathcal{U}$ with associated entropy solutions u and $\tilde{u}$.

Given $t \geq 0$, Proposition 3.21 yields again the convergence of $[S_h u^0]_x(t,.)$ to $u_x(t,.)$ in the weak sense of probability measures. Since this holds also for $\tilde{u}^0$, we obtain

$$W_p(u_x(t,.), \tilde{u}_x(t,.)) \leq \liminf_{h\to 0} W_p([S_h u^0]_x(t,.), [S_h \tilde{u}^0]_x(t,.))$$

by Proposition 3.15. But, for each h,

$$W_p([S_h u^0]_x(t,.), [S_h \tilde{u}^0]_x(t,.)) \leq W_p(u^0_x, \tilde{u}^0_x)$$

by Proposition 3.19, so finally

$$W_p(u_x(t,.), \tilde{u}_x(t,.)) \leq W_p(u^0_x, \tilde{u}^0_x).$$

This concludes the argument. □

3.5.2 Proof of Corollary 3.9

We recall that in the introduction we have defined a distance on each $\mathcal{U}_p$ by letting

$$d_p(u, \tilde{u}) \;=\; W_p(u_x, \tilde{u}_x),$$

and we now prove that, given $p \geq 1$ and $u^0 \in \mathcal{U}_p$, the entropy solution u to the conservation law (3.1) belongs to $\mathcal{C}([0, +\infty[, \mathcal{U}_p)$.

We first note, in view of the proof of Proposition 3.26, that $u(t, .)$ indeed belongs to $\mathcal{U}_p$ for all $t \geq 0$.

Then, given $s \geq 0$, we need to prove that $d_p(u(t, .), u(s, .))\,(= W_p(u_x(t, .), u_x(s, .)))$ tends to 0 as t goes to s. Indeed, on one hand $u(t, .)$ tends to $u(s, .)$ in $L^1_{loc}(\mathbb{R})$ by Theorem 3.1, so $u_x(t, .)$ tends to $u_x(s, .)$, first in the distribution sense, then in the weak sense of probability measures.

On the other hand, given $T > s$, we now prove that $\sup\limits_{0 \leq t \leq T} \displaystyle\int_{|x| \geq R} |x|^p \, du_x(t, x)$ goes to 0 as R goes to infinity. For this, given $\varepsilon > 0$, let R such that

$$\sup_{0 \leq h \leq T} \sup_{0 \leq t \leq T} \int_{|x| \geq R} |x|^p \, d[S_h u^0]_x(t, x) \leq \varepsilon$$

by Proposition 3.25. Let then $\varphi \in \mathcal{C}^\infty_c(\mathbb{R})$ such that $0 \leq \varphi \leq 1$ and $\varphi(x) = 0$ if $|x| \leq R$.

On one hand

$$\int_{\mathbb{R}} \varphi(x)|x|^p \, d[S_h u^0]_x(t, x) \;\to\; \int_{\mathbb{R}} \varphi(x)|x|^p \, du_x(t, x)$$

as h goes to 0 since $\varphi(x)|x|^p \in \mathcal{C}^\infty_c(\mathbb{R})$ and $[S_h u^0]_x(t, .)$ tends to $u_x(t, .)$ in distribution sense. On the other hand

$$\int_{\mathbb{R}} \varphi(x)|x|^p \, d[S_h u^0]_x(t, x) \;\leq\; \varepsilon$$

for all $0 \leq h, t \leq T$. Hence at the limit

$$\int_{\mathbb{R}} \varphi(x)|x|^p \, du_x(t, x) \;\leq\; \varepsilon$$

for all $t \leq T$, from which it follows that

$$\sup_{0 \leq t \leq T} \int_{|x| \geq R} |x|^p \, du_x(t, x) \;\leq\; \varepsilon,$$

which means that indeed $\sup\limits_{0 \leq t \leq T} \displaystyle\int_{|x| \geq R} |x|^p \, du_x(t, x)$ goes to 0 as R goes to infinity.

From these two results we deduce the continuity result by Proposition 3.14. $\square$

3.5.3 Proof of Corollary 3.10

Given $t \geq 0$, $S_h u^0(t,.)$ converges to $u(t,.)$ in $L^1_{loc}(\mathbb{R})$ by Proposition 3.21, so for all $s, t, n \geq 0$ we have

$$\|u(t,.) - u(s,.)\|_{L^1(]-n,n[)} = \lim_{h\to 0} \|S_h u^0(t,.) - S_h u^0(s,.)\|_{L^1(]-n,n[)}.$$

But

$$\|S_h u^0(t,.) - S_h u^0(s,.)\|_{L^1(]-n,n[)} \leq \|S_h u^0(t,.) - S_h u^0(s,.)\|_{L^1(\mathbb{R})} \leq |t-s|\, \|f'\|_{L^\infty(\mathbb{R})}$$

for all $h \geq 0$ by Proposition 3.20, so letting h go to 0 we get

$$\|u(t,.) - u(s,.)\|_{L^1(]-n,n[)} \leq |t-s|\, \|f'\|_{L^\infty(\mathbb{R})}.$$

Since this holds for all $n \geq 0$, we obtain Corollary 3.10. □

3.6 Extension to viscous conservation laws

In this section we let ν be some positive number and f be some locally Lipschitz real-valued function on $\mathbb{R}$, and we consider the viscous conservation law

$$\begin{cases} u_t + f(u)_x = \nu\, u_{xx} & t > 0,\ x \in \mathbb{R} \\ u(0,.) \;=\; u^0, \end{cases} \tag{3.15}$$

with unknown $u = u(t,x) \in \mathbb{R}$ and initial datum $u^0 \in L^\infty(\mathbb{R})$.

We shall prove that the W_p contraction property given in Theorem 3.6 in the inviscid case when $\nu = 0$ also holds in the viscous case when ν is positive.

In this section we consider solutions in the sense of distributions : more precisely a function u in $L^\infty(]0,+\infty[\times\mathbb{R})$ is a solution of (3.15) if

$$\int_0^{+\infty}\!\!\int_{\mathbb{R}} \big(u\,\varphi_t + f(u)\,\varphi_x + \nu\, u\,\varphi_{xx}\big)\, dt\, dx \;+\; \int_{\mathbb{R}} u^0(x)\,\varphi(0,x)\, dx \;= 0$$

for any φ in $\mathcal{C}^\infty_c([0,+\infty[\times\mathbb{R})$. It is known (see [102] for instance) that, given $u^0 \in L^\infty(\mathbb{R})$, there exists a unique solution to (3.15) in $L^\infty(]0,+\infty[\times\mathbb{R})$.

The set $\mathcal{U}$ is preserved by the viscous conservation law in the sense of

Proposition 3.27. *Any solution u to* (3.15) *with initial datum $u^0 \in \mathcal{U}$ is such that $u(t,.)$ belong to $\mathcal{U}$ for all $t \geq 0$.*

Proof. Given $t \geq 0$ the solution $u(t,.)$ takes values between 0 and 1 by maximum principle, has total variation bounded by the total variation of u^0, which is 1, and is almost everywhere nondecreasing. In particular it is continuous but at countably many points and has limits ≥ 0 and ≤ 1 at $-\infty$ and $+\infty$ respectively. These limits are actually equal to 0 and 1 respectively since the quantity $\displaystyle\int_{\mathbb{R}} du_x(t,x)$ (which is equal to the total variation of $u(t,.)$) is conserved, and equals 1 (see for instance [101]). $\square$

As in the case of the inviscid conservation law discussed above we shall prove

Theorem 3.28. *Given a locally Lipschitz real-valued function f on $\mathbb{R}$ and two initial data u^0 and $\tilde{u}^0$ in $\mathcal{U}$, let u and $\tilde{u}$ be the associated solutions to* (3.15)*. Then, for any $t \geq 0$ and $p \geq 1$, we have (with possibly infinite values)*

$$W_p(u_x(t,.), \tilde{u}_x(t,.)) \;\leq\; W_p(u_x^0, \tilde{u}_x^0).$$

We briefly mention how this contraction property for the viscous conservation law enables to recover the same property for the inviscid equation, given in Theorem 3.6. Given some initial datum u^0 in $\mathcal{U}$ and $\nu > 0$, let indeed u_ν be the corresponding solution to the viscous equation (3.15). Then it is known (see [101] for instance) that $u_\nu(t,.)$ converges in $L^1_{loc}(\mathbb{R})$ to the solution $u(t,.)$ to the inviscid conservation law (3.1) with initial datum u^0. From this the argument already used in Section 3.5.1 (with $S_h u^0(t,.)$ intead of $u_\nu(t,.)$) enables to recover Theorem 3.6.

The proof of Theorem 3.28 is an extension of the argument given in the inviscid case in Sections 3.4 and 3.5, and goes in several steps. The property is first proved for a time-discretized solution of equation (3.15), then for the true solution to equation (3.15) by using the convergence of the approximate solution.

3.6.1 Time-discretization of the equation

In this section we define a time-discretized solution of the equation by using the discretization scheme of the inviscid conservation law discussed in Section 3.4.

The procedure goes in the following two steps. Given some time step $h > 0$, we first map v to $T_h v$ as in Section 3.4; then we let $T_h v$ evolve according to the heat equation with viscosity ν on a time length h, that is, map it to

$$\mathcal{T}_h v = K_h * T_h v$$

where K_h is the heat kernel defined on $\mathbb{R}$ by

$$K_h(z) = \frac{1}{\sqrt{4\pi h\nu}}\, e^{-\frac{z^2}{4h\nu}}.$$

Letting g be the standard Gaussian on $\mathbb{R}$ defined by

$$g(y) \;=\; \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}},$$

and recalling that T_h satisfies

$$T_h v(x) = \int_0^1 jv(x - hf'(w), w)\, dw$$

where

$$jv(x,w) = \begin{cases} 1 & \text{if } v(x) > w \\ 0 & \text{if } v(x) \le w, \end{cases}$$

we have

$$\begin{aligned} \mathcal{T}_h v(x) &= \int_{\mathbb{R}} T_h v(x-z)\, K_h(z)\, dz = \int_{\mathbb{R}} T_h(x - \sqrt{2h\nu}y)\, g(y)\, dy \\ &= \int_{\mathbb{R}} \int_0^1 jv(x - \sqrt{2h\nu}y - hf'(w), w)\, g(y)\, dy\, dw. \end{aligned}$$

In particular we recover in our case the time discretization introduced in [26].

From the properties satisfied by T_h can be deduced

Proposition 3.29. *Let h be some positive number and $\mathcal{T}_h$ defined as above. Then*
i) $\mathcal{T}_h v$ belongs to $\mathcal{U}$ (resp. $\mathcal{U}_p$) for any v in $\mathcal{U}$ (resp. $\mathcal{U}_p$).
ii) For any v and $\tilde{v}$ in $\mathcal{U}$ and $p \ge 1$ we have (with possibly infinite values)

$$W_p([\mathcal{T}_h v]_x, [\mathcal{T}_h \tilde{v}]_x) \;\le\; W_p(v_x, \tilde{v}_x).$$

Proof. i) Let v be given in $\mathcal{U}$. Then $T_h v$ belongs to $\mathcal{U}$ by Definition 3.16, and thus so does $\mathcal{T}_h v$ as the convolution of $T_h v$ with a probability measure.

If moreover v belongs to $\mathcal{U}_p$, that is, has finite moment of order p, then so does $T_h v$ by Proposition 3.17. Since so does the Gaussian K_h also, this ensures that

$$\begin{aligned} \int_{\mathbb{R}} |x|^p [\mathcal{T}_h v]_x(x)\, dx &= \iint_{\mathbb{R}^2} |x+y|^p K_h(y) [T_h v]_x(x)\, dx\, dy \\ &\le 2^{p-1} \Big(\int_{\mathbb{R}} |y|^p K_h(y)\, dy + \int_{\mathbb{R}} |x|^p [T_h v]_x(x)\, dx \Big) \end{aligned}$$

is finite. Hence $\mathcal{T}_h v$ also belongs to $\mathcal{U}_p$.

ii) K_h being a probability measure, and W_p^p being convex, we have

$$W_p([\mathcal{T}_h v]_x, [\mathcal{T}_h \tilde{v}]_x) \;\le\; W_p([T_h v]_x, [T_h \tilde{v}]_x),$$

which in turn is bounded by $W_p(v_x, \tilde{v}_x)$ by Proposition 3.17 again. □

Then we define an approximate solution $\mathcal{S}_h u^0$ to (3.15) by iterating the $\mathcal{T}_h$ operator according to

Definition 3.30. Let h be some positive number and $v \in \mathcal{U}$. For any $t \ge 0$ decomposed as $t = (N+n)h$ with $N \in \mathbb{N}$ and $0 \le n < 1$, we let

$$\mathcal{S}_h v(t,.) \;=\; (1-n)\, \mathcal{T}_h^N v(.) \;+\; n\, \mathcal{T}_h^{N+1} v(.)$$

where $\mathcal{T}_h^0 v = v$ and $\mathcal{T}_h^{N+1} v = \mathcal{T}_h(\mathcal{T}_h^N v)$.

From Proposition 3.29 we deduce the following W_p contraction property :

Proposition 3.31. *Let h be some positive number and $\mathcal{S}_h$ defined as above. Then*
i) $\mathcal{S}_h v$ belongs to $\mathcal{U}$ (resp. $\mathcal{U}_p$) for any v in $\mathcal{U}$ (resp. $\mathcal{U}_p$).
ii) For any v and $\tilde{v}$ in $\mathcal{U}$, $p \geq 1$ and $t \geq 0$, we have (with possibly infinite values)

$$W_p([\mathcal{S}_h v]_x(t,.), [\mathcal{S}_h \tilde{v}]_x(t,.)) \;\leq\; W_p(v_x, \tilde{v}_x).$$

The following $L^1(\mathbb{R})$ contraction property shall be useful to prove the convergence of the discretized solution $\mathcal{S}_h u^0$ toward the solution to (3.15) :

Proposition 3.32. *Let h be some fixed positive number and S_h defined as above. Then, for any v in $\mathcal{U}$, twice derivable with v'' in $L^1(\mathbb{R})$, and any $s, t \geq 0$, we have*

$$\|\mathcal{S}_h v(t,.) - \mathcal{S}_h v(s,.)\|_{L^1(\mathbb{R})} \;\leq\; |t-s| \,\big[\|f'\|_{L^\infty(]0,1[)} + \nu\, \|v''\|_{L^1(\mathbb{R})}\big].$$

Proof. The argument goes in several steps.
1. Let u, v in $\mathcal{U}$ with $v - u \in L^1(\mathbb{R})$. Then

$$\|\mathcal{T}_h v - \mathcal{T}_h u\|_{L^1(\mathbb{R})} \leq \; \|v-u\|_{L^1(\mathbb{R})}$$

whence

$$\|\mathcal{T}_h^k v - \mathcal{T}_h^k u\|_{L^1(\mathbb{R})} \leq \|v-u\|_{L^1(\mathbb{R})} \tag{3.16}$$

for any integer number k.
Indeed

$$\|\mathcal{T}_h v - \mathcal{T}_h u\|_{L^1(\mathbb{R})} = \|T_h * K_h v - T_h * K_h u\|_{L^1(\mathbb{R})} \leq \|T_h v - T_h u\|_{L^1(\mathbb{R})}\, \|K_h\|_{L^1(\mathbb{R})}$$

where $\|K_h\|_{L^1(\mathbb{R})} = 1$ and

$$\begin{aligned} \|T_h v - T_h u\|_{L^1(\mathbb{R})} &= \int_{\mathbb{R}} \Big| \int_0^1 jv(x - hf'(w), w) - ju(x-hf'(w), w)\, dw \Big|\, dx \\ &\leq \int_0^1 \int_{\mathbb{R}} \big|jv(y,w) - ju(y,w)\big|\, dy\, dw \\ &= \|v-u\|_{L^1(\mathbb{R})} \end{aligned}$$

according to the proof of Proposition 3.13.

2. Let v in $\mathcal{U}$ with second derivative v'' in $L^1(\mathbb{R})$. Then

$$\|K_h * v - v\|_{L^1(\mathbb{R})} \leq h\, \nu\, \|v''\|_{L^1(\mathbb{R})}.$$

Indeed the relation $2\,h\,\nu\,K_h'(y) = -y\,K_h(y)$ and two integrations by parts ensure that, given $\phi \in \mathcal{C}_c^2(\mathbb{R})$,

$$\begin{aligned}
\int_{\mathbb{R}} [K_h * v(x) - v(x)]\,\phi(x)\,dx &= \iint_{\mathbb{R}^2} v(x)\,K_h(y)\,[\phi(x+y) - \phi(x)]\,dy\,dx \\
&= \iint_{\mathbb{R}^2} v(x)\,K_h(y)\Big(\int_0^1 \phi'(x+sy)\,y\,ds\Big)\,dx\,dy \\
&= \int_0^1 \int_{\mathbb{R}} v(x)\Big(\int_{\mathbb{R}} (-2\,h\,\nu) K_h'(y)\phi'(x+sy)\,dy\Big)\,dx\,ds \\
&= 2\,h\,\nu \int_0^1 \int_{\mathbb{R}} v(x)\Big(\int_{\mathbb{R}} K_h(y)\phi''(x+sy)\,s\,dy\Big)\,ds\,dx \\
&= 2\,h\,\nu \int_0^1 s \int_{\mathbb{R}} K_h(y)\Big(\int_{\mathbb{R}} v''(x)\phi(x+sy)\,dx\Big)\,ds\,dy \\
&\leq h\,\nu\,\|v''\|_{L^1(\mathbb{R})}\,\|\phi\|_{L^\infty(\mathbb{R})}.
\end{aligned}$$

3. By triangular inequality, the bound (3.13) and finally step 2, we obtain

$$\begin{aligned}
\|\mathcal{T}_h v - v\|_{L^1(\mathbb{R})} &\leq \|K_h * T_h v - K_h * v\|_{L^1(\mathbb{R})} + \|K_h * v - v\|_{L^1(\mathbb{R})} \\
&\leq \|T_h v - v\|_{L^1(\mathbb{R})}\,\|K_h\|_{L^1(\mathbb{R})} + \|K_h * v - v\|_{L^1(\mathbb{R})} \\
&\leq h\,\big[\|f'\|_{L^\infty(]0,1[)} + \nu\,\|v''\|_{L^1(\mathbb{R})}\big]. \qquad (3.17)
\end{aligned}$$

4. We can now conclude the argument. For this purpose we decompose s as $(M+m)h$ and t as $(N+n)h$ with $M, N \in \mathbb{N}$ and $0 \leq m, n < 1$. For instance assuming that $N > M$, we obtain

$$\begin{aligned}
\|\mathcal{S}_h v(t,.) - \mathcal{S}_h v(s,.)\|_{L^1(\mathbb{R})} &\leq n\,\|\mathcal{T}_h^N(\mathcal{T}_h v) - \mathcal{T}_h^N v\|_{L^1(\mathbb{R})} + \sum_{k=M+1}^{N-1} \|\mathcal{T}_h^k(\mathcal{T}_h v) - \mathcal{T}_h^k v\|_{L^1(\mathbb{R})} \\
&\quad + (1-m)\,\|\mathcal{T}_h^M(\mathcal{T}_h v) - \mathcal{T}_h^M v\|_{L^1(\mathbb{R})} \\
&\leq \Big(n + \sum_{k=M+1}^{N-1} 1\ + 1 - m\Big)\|\mathcal{T}_h v - v\|_{L^1(\mathbb{R})} \\
&\leq |t-s|[\|f'\|_{L^\infty(]0,1[)} + \nu\,\|v''\|_{L^1(\mathbb{R})}]
\end{aligned}$$

by (3.16) and (3.17) successively. □

3.6.2 Convergence of the scheme

In this section we prove

Proposition 3.33. *Let u^0 be in $\mathcal{U}$. Then, as h goes to 0, the family $(\mathcal{S}_h u^0)_h$ converges in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$ toward the solution to (3.15) with initial datum u^0.*

The proof is similar to the one given in Section 3.4 in the inviscid case, going in the following steps.

1. Compactness of the scheme

It is given by

Proposition 3.34. *Given v in $\mathcal{U}$, the family $(\mathcal{S}_h v)_h$ is relatively compact in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$.*

Proof. 1. Let $\mathcal{V}$ be the subset of $\mathcal{U}$ composed of the twice derivable functions v with v'' in $L^1(\mathbb{R})$. Then the conclusion of the proposition holds for any v in $\mathcal{V}$ by the argument of Proposition 3.22, in view of Proposition 3.32.

2. Let now v be any element in $\mathcal{U}$ and $(\mathcal{S}_{h_m} v)_m$ be a sequence of the family $(\mathcal{S}_h v)_h$, where $(h_m)_m$ is a sequence converging to 0 as m goes to infinity.

Let also $(v_n)_n$ be a sequence in $\mathcal{V}$ such that $\|v - v_n\|_{L^1(\mathbb{R})}$ tend to 0 as n goes to infinity. For instance, one may let $v_n = \rho_n * v$ where $\rho_n(x) = n\,\rho(nx)$ for some nonnegative $\mathcal{C}^\infty_c(\mathbb{R})$ function ρ with unit integral; then one can check that

$$\|v - v_n\|_{L^1(\mathbb{R})} \leq \frac{1}{n} \int_{\mathbb{R}} |z|\, \rho(z)\, dz. \tag{3.18}$$

For any n, the function v_n belongs to $\mathcal{V}$ so by the first step the sequence $(\mathcal{S}_{h_m} v_n)_m$ is relatively compact in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$. Then, by a diagonal argument, there exist a sequence $(m_k)_k$ tending to infinity, that we shall denote $(m)_m$ again, and a sequence $(w_n)_n$ in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$ such that, for any n, $\mathcal{S}_{h_m} v_n$ converge to w_n in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$ as m goes to infinity.

Let us note that $(w_n)_n$ is a Cauchy sequence in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$. Indeed it follows from (3.16) that

$$\sup_m \sup_t \|\mathcal{S}_{h_m} v_n(t,.) - \mathcal{S}_{h_m} v_p(t,.)\|_{L^1(\mathbb{R})} \leq \|v_n - v_p\|_{L^1(\mathbb{R})}$$

for any n and p. In particular at the limit in m

$$\sup_t \|w_n(t,.) - w_p(t,.)\|_{L^1(\mathbb{R})} \leq \|v_n - v_p\|_{L^1(\mathbb{R})}, \tag{3.19}$$

so that $(w_n)_n$ is a Cauchy sequence in the complete space $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$ by (3.18). Let w_∞ be its limit.

Then, for any T, R, m and n we have

$$\begin{aligned}
\sup_{0\leq t\leq T} \|\mathcal{S}_{h_m} v(t,.) - w_\infty(t,.)\|_{L^1(]-R,+R[)} &\leq \sup_{0\leq t\leq T} \|\mathcal{S}_{h_m} v(t,.) - \mathcal{S}_{h_m} v_n(t,.)\|_{L^1(]-R,+R[)} \\
&+ \sup_{0\leq t\leq T} \|\mathcal{S}_{h_m} v_n(t,.) - w_n(t,.)\|_{L^1(]-R,+R[)} + \sup_{0\leq t\leq T} \|w_n(t,.) - w_\infty(t,.)\|_{L^1(]-R,+R[)} \\
\leq \|v - v_n\|_{L^1(]-R,+R[)} &+ \sup_{0\leq t\leq T} \|\mathcal{S}_{h_m} v_n(t,.) - w_n(t,.)\|_{L^1(]-R,+R[)} + \sup_{0\leq t\leq T} \|w_n(t,.) - w_\infty\|_{L^1(]-R,+R[)}.
\end{aligned}$$

Given $\varepsilon > 0$, there exists n such that the first and third terms be bounded by ε, and for this n there exists some M such that the second term also be bounded by ε for any $m \geq M$.

Since R and T are arbitrary, this means that the original sequence $(\mathcal{S}_{h_m}v)_m$ of the family $(\mathcal{S}_h v)_h$ has a subsequence which converges in $C([0,+\infty[, L^1_{loc}(\mathbb{R}))$ (toward w_∞). This concludes the argument. $\square$

2. An approximate solution

Proposition 3.35. *Let v be in $\mathcal{U}$ and ϕ in $\mathcal{C}^\infty_c(\mathbb{R})$ with support included in $[-R,R]$. Then, in the above notation, we have*

$$\left|\int_{\mathbb{R}}[\mathcal{T}_h v(x)-v(x)]\phi(x)\,dx - h\,\nu\int_{\mathbb{R}} v(x)\phi''(x)\,dx - h\int_{\mathbb{R}} f(v(x))\phi'(x)\,dx\right| \\ \leq \left[h^2(M^2(R+hM)+\nu\,M)+\frac{4}{3\sqrt{\pi}}(h\nu)^{3/2}\right]\|\phi''\|_{L^\infty(\mathbb{R})}$$

where $M=\|f'\|_{L^\infty(]0,1[)}$.

Proof. By definition of $\mathcal{T}_h$ and triangular inequality we have

$$\begin{aligned}&\left|\int_{\mathbb{R}}[\mathcal{T}_h v(x)-v(x)]\phi(x)\,dx - h\,\nu\int_{\mathbb{R}} v(x)\phi''(x)\,dx - h\int_{\mathbb{R}} f(v(x))\phi'(x)\,dx\right| \\ &\leq \left|\int_{\mathbb{R}}[K_h * V(x)-V(x)]\phi(x)\,dx - h\,\nu\int_{\mathbb{R}} V(x)\phi''(x)\,dx\right| + h\,\nu\left|\int_{\mathbb{R}}[T_h v(x)-v(x)]\phi''(x)\,dx\right| \\ &\qquad + \left|\int_{\mathbb{R}}[T_h v(x)-v(x)]\phi(x)\,dx - h\int_{\mathbb{R}} f(v(x))\phi'(x)\,dx\right| \quad (3.20)\end{aligned}$$

where in the first term we have let $V=T_h v$.

1. The first term in (3.20) is bounded by $\dfrac{4}{3\sqrt{\pi}}(h\nu)^{3/2}\|\phi''\|_{L^\infty(\mathbb{R})}$. Indeed

$$\begin{aligned}\int_{\mathbb{R}}[K_h * V(x)-V(x)]\phi(x)\,dx &= \iint_{\mathbb{R}^2} V(x)[\phi(x+y)-\phi(x)]\,K_h(y)\,dx\,dy \\ &= \iint_{\mathbb{R}^2} V(x)\,\phi'(x)\,y\,K_h(y)\,dx\,dy + \iint_{\mathbb{R}^2} V(x)\,\phi''(x)\,\frac{y^2}{2}\,K_h(y)\,dx\,dy \\ &\quad + \iint_{\mathbb{R}^2} V(x)\Big(\int_0^1 \phi^{(3)}(x+ty)\,(1-t)^2\,dt\Big)\frac{y^3}{2}\,K_h(y)\,dx\,dy\end{aligned}$$

by Taylor's formula.

The first term in the right-hand side is equal to 0 since K_h is an even function. The second term is equal to $h\,\nu\displaystyle\int_{\mathbb{R}} V(x)\phi''(x)\,dx$ since $\displaystyle\int_{\mathbb{R}} y^2\,K_h(y)\,dy = 2\,h\,\nu$. Finally the third term is equal to

$$-\int_{\mathbb{R}}\int_0^1 \frac{y^3}{2}\,K_h(y)(1-t)^2\,dt\,\Big(\int_{\mathbb{R}}\phi''(x+ty)\,dV_x(x)\Big)\,dy\,dt$$

where V_x is a probability measure on $\mathbb{R}$, so is bounded by

$$\int_{\mathbb{R}} \frac{|y|^3}{2}\, K_h(y)\, dy \int_0^1 (1-t)^2\, \|\phi''\|_{L^\infty(\mathbb{R})} = \frac{4}{3\sqrt{\pi}}(h\nu)^{3/2}\|\phi''\|_{L^\infty(\mathbb{R})}.$$

2. The second term in (3.20) is bounded by

$$h\,\nu\|T_h v - v\|_{L^1(\mathbb{R})}\, \|\phi''\|_{L^\infty(\mathbb{R})} \leq h^2\,\nu\, M\, \|\phi''\|_{L^\infty(\mathbb{R})}$$

according to (3.13).

3. The third term in (3.20) is bounded by $h^2\, \|\phi''\|_{L^\infty(\mathbb{R})} M^2(R + h\, M)$ according to Proposition 3.23 with $E(x) = x$ and $E(x) = -x$ (for which $F(x) = f(x)$ and $F(x) = -f(x)$ respectively).

This concludes the argument. □

3. Proof of the convergence of the scheme

In this section we prove Proposition 3.33 by adapting the argument of Proposition 3.21 to the viscous case, where solutions are meant in the sense of distributions.

For this purpose, given u^0 in $\mathcal{U}$, we consider a sequence of the family $(S_h u^0)_h$ converging to a function u in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$ by Proposition 3.34 ; we shall prove by Proposition 3.35 that u is a solution to the viscous conservation law (3.15) with initial datum u^0, that is, that u belongs to $L^\infty(]0,+\infty[\times\mathbb{R})$ and satisfies

$$\int_0^{+\infty}\!\!\int_{\mathbb{R}} \big(u\,\varphi_t + f(u)\,\varphi_x + \nu\, u\,\varphi_{xx}\big)\, dt\, dx \,+ \int_{\mathbb{R}} u^0(x)\,\varphi(0,x)\, dx \,= 0$$

for any φ in $\mathcal{C}^\infty_c([0,+\infty[\times\mathbb{R})$.

First of all the converging sequence, that we shall denote $(S_h u^0)_h$ again, is bounded in $L^\infty(]0,+\infty[\times\mathbb{R})$, so that its limit u also belongs to $L^\infty(]0,+\infty[\times\mathbb{R})$.

Then we let $\varphi \in \mathcal{C}^\infty_c([0,+\infty[\times\mathbb{R})$ and then $T \geq 0$ be such that $\varphi(t,x) \equiv 0$ for $t \geq T$. Given $h \geq 0$ we let N stand for $[T/h]$. We shall also let C denote various constants depending on φ (hence on T), $\|f'\|_{L^\infty(]0,1[)}$ and ν, but neither on k nor h. Applying Proposition 3.35 to $v = \mathcal{S}_h u^0(kh,.)$ and $\phi = \varphi(kh,.)$ for $k = 0,\ldots,N$ ensures that

$$\int_{\mathbb{R}} \big[\mathcal{S}_h u^0((k+1)h,x) - \mathcal{S}_h u^0(kh,x)\big]\,\varphi(kh,x)\, dx \\ -\, h\,\nu \int_{\mathbb{R}} \mathcal{S}_h u^0(kh,x)\,\varphi_{xx}(kh,x)\, dx - h\int_{\mathbb{R}} f\big(\mathcal{S}_h u^0(kh,x)\big)\,\varphi_x(kh,x)\, dx$$

is bounded by $Ch^{3/2}$ for $k = 0,\ldots,N$. Hence, summing on $k = 0,\ldots,N$ and making a change of indexes,

$$
\begin{aligned}
-\sum_{k=1}^{N}\int_{\mathbb{R}} \mathcal{S}_h u^0(kh,x)\ [\varphi(kh,x)-\varphi((k-1)h,x)]\ dx \\
-h\,\nu\sum_{k=0}^{N}\int_{\mathbb{R}} \mathcal{S}_h u^0(kh,x)\,\varphi_{xx}(kh,x)\,dx - h\sum_{k=0}^{N}\int_{\mathbb{R}} f\big(\mathcal{S}_h u^0(kh,x)\big)\,\varphi_x(kh,x)\,dx \\
-\int_{\mathbb{R}} u^0(x)\,\varphi(0,x)\,dx - \int_{\mathbb{R}} S_h u^0((N+1)h,x)\,\varphi(Nh,x)\,dx
\end{aligned}
\tag{3.21}
$$

is bounded by $C\sqrt{h}$.

Then, by Proposition 3.32 and the convergence of $(\mathcal{S}_h u^0)_h$ towards u in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$, we can follow the argument of Proposition 3.21 to check that the first sum in (3.21) converges to $\displaystyle\int_0^T\int_{\mathbb{R}} u(t,x)\,\varphi_t(t,x)\,dt\,dx$, the second sum to $\nu\displaystyle\int_0^T\int_{\mathbb{R}} u(t,x)\,\varphi_{xx}(t,x)\,dt\,dx$, and the third one to $\displaystyle\int_0^T\int_{\mathbb{R}} f(u(t,x))\,\varphi_x(t,x)\,dt\,dx$. Since finally the last term in (3.21) tends to 0 as h goes to 0, it follows that u satisfies

$$
\int_0^{+\infty}\!\!\int_{\mathbb{R}}\big(u\,\varphi_t + f(u)\,\varphi_x + \nu\,u\,\varphi_{xx}\big)\,dt\,dx \;+\; \int_{\mathbb{R}} u^0(x)\,\varphi(0,x)\,dx \;=0.
$$

Consequently any sequence of $(\mathcal{S}_h u^0)_h$ has a subsequence converging in $\mathcal{C}([0,+\infty[, L^1_{loc}(\mathbb{R}))$ to a solution to (3.15) with initial datum u^0. This concludes the proof of Proposition 3.33 by uniqueness of this solution. □

3.6.3 Proof of Theorem 3.28

In this section we let p be some real number ≥ 1 and we give the proof of Theorem 3.28.

We let u^0 and $\tilde{u}^0$ be two initial data in $\mathcal{U}$ and we denote u and $\tilde{u}$ their respective solutions to (3.15).

Given $t \geq 0$, $\mathcal{S}_h u^0(t,.)$ converges to $u(t,.)$ in $L^1_{loc}(\mathbb{R})$ by Proposition 3.33, so $[\mathcal{S}_h u^0]_x(t,.)$ converges to $u_x(t,.)$ first in the distribution sense, then in the weak sense of probability measures. Since this holds for $\tilde{u}^0$ also we have

$$
W_p(u_x(t,.),\tilde{u}_x(t,.)) \leq \liminf_{h\to 0} W_p([\mathcal{S}_h u^0]_x(t,.),[\mathcal{S}_h \tilde{u}^0]_x(t,.))
$$

by Proposition 3.15. But, for each h,

$$
W_p([\mathcal{S}_h u^0]_x(t,.),[\mathcal{S}_h \tilde{u}^0]_x(t,.)) \leq W_p(u^0_x,\tilde{u}^0_x)
$$

by Proposition 3.31, so finally

$$
W_p(u_x(t,.),\tilde{u}_x(t,.)) \leq W_p(u^0_x,\tilde{u}^0_x).
$$

This concludes the argument. □

3.7 Extension to conservation laws with different flux functions

We have seen in the previous sections that if ν is a nonnegative number and f is a locally Lipschitz function, then any two solutions u and $\tilde{u}$ to the scalar conservation law

$$u_t + f(u)_x = \nu\, u_{xx}, \qquad t > 0, \quad x \in \mathbb{R}$$

with respective initial data u^0 and $\tilde{u}^0$ in $\mathcal{U}$ satisfy the contraction property

$$W_p(u_x(t,.), \tilde{u}_x(t,.)) \leq W_p(u^0_x, \tilde{u}^0_x)$$

for any $t \geq 0$ and $p \geq 1$. In the whole section, by solution we mean entropy solution in the inviscid case when $\nu = 0$, and solution in the sense of distributions in the viscous case when $\nu > 0$.

In this section we give an extension of this result for solutions u and $\tilde{u}$ to conservation laws with different flux functions.

More precisely we let again ν be some nonnegative number, and f and $\tilde{f}$ be two locally Lipschitz functions on $\mathbb{R}$. Then we let u be the solution to

$$u_t + f(u)_x = \nu\, u_{xx}, \qquad t > 0, \quad x \in \mathbb{R} \tag{3.22}$$

with initial datum u^0 in $\mathcal{U}$, and $\tilde{u}$ be the solution to

$$\tilde{u}_t + \tilde{f}(\tilde{u})_x = \nu\, \tilde{u}_{xx}, \qquad t > 0, \quad x \in \mathbb{R}$$

with initial datum $\tilde{u}^0$ in $\mathcal{U}$. We shall prove

Theorem 3.36. *In the above notation we have, with possibly infinite values,*

$$W_p(u_x(t,.), \tilde{u}_x(t,.)) \leq W_p(u^0_x, \tilde{u}^0_x) + t\, \|f' - \tilde{f}'\|_{L^p(]0,1[)} \tag{3.23}$$

for any $t \geq 0$ and $p \geq 1$.

In particular we recover Theorems 3.6 and 3.7 for $\tilde{f} = f$. On the other hand, relation (3.23) also writes

$$W_p(u_x(t,.), \tilde{u}_x(t,.)) \leq W_p(u_x(s,.), \tilde{u}_x(t,.)) + (t - s)\, \|f' - \tilde{f}'\|_{L^p(]0,1[)}$$

for any $t \geq s \geq 0$; in this form it extends the result obtained by Y. Brenier [29, Section 5] in the case when $\nu = 0$ and $p = 2$.

Proof. We slightly change the notation introduced in Sections 3.4 and 3.6 to define the approximate solution to (3.22). More precisely we add an exponent f (or $\tilde{f}$) to mean that the evolution is governed by the flux function f (or $\tilde{f}$); for instance equation (3.22) shall be discretized by means of the T^f_h function defined as follows (see Definition 3.16) :

Definition 3.37. Let v in $\mathcal{U}$ with generalized inverse $X(0,.)$. Then, given $h>0$, and letting

$$X^f(h,w) = X(0,w) + h\, f'(w)$$

for almost every $w \in]0,1[$, we define the $\mathcal{U}$ function $T_h^f v$ on $\mathbb{R}$ by

$$T_h^f v(x) = \int_0^1 \mathbf{1}_{\{X^f(h,w)\le x\}}(w)\, dw.$$

To use common notation in the case when $\nu = 0$ and $\nu > 0$, we let K_h be again the heat kernel defined on $\mathbb{R}$ by

$$K_h(z) = \frac{1}{\sqrt{4\pi h\nu}}\, e^{-\frac{z^2}{4h\nu}}$$

if $\nu > 0$, and K_h be the Dirac mass δ_0 at 0 if $\nu = 0$ (which by the way is the weak limit of the heat kernel above as ν goes to 0). Then, as in Section 3.6, we let

$$\mathcal{T}_h^f = K_h * T_h^f,$$

so that $\mathcal{T}_h^f = \delta_0 * T_h^f = T_h^f$ again in the inviscid case, and we define an approximate solution $\mathcal{S}_h^f u^0$ to (3.22) by iterating the $\mathcal{T}_h^f$ operator as in Definition 3.30.

We use similar notation for $\tilde{f}$.

Now the proof goes in the following 4 steps.

1. Let v be given in $\mathcal{U}$. Then

$$W_p([T_h^f v]_x, [T_h^{\tilde{f}} v]_x) \le h\, \|f' - \tilde{f}'\|_{L^p(]0,1[)}.$$

Indeed, if $X(0,.)$ is the generalized inverse of v, then $[T_h^f v]_x$ is the law of

$$X^f(h,w) = X(0,w) + h\, f'(w)$$

whereas $[T_h^{\tilde{f}} v]_x$ is the law of

$$X^{\tilde{f}}(h,w) = X(0,w) + h\, \tilde{f}'(w).$$

In particular

$$W_p^p([T_h^f v]_x, [T_h^{\tilde{f}} v]_x) \le \int_0^1 |X^f(h,w) - X^{\tilde{f}}(h,w)|^p\, dw = h^p \int_0^1 |f'(w) - \tilde{f}'(w)|^p\, dw.$$

2. Let v and $\tilde{v}$ be given in $\mathcal{U}$. Then

$$W_p([\mathcal{T}_h^f v]_x, [\mathcal{T}_h^{\tilde{f}} \tilde{v}]_x) \le W_p(v_x, \tilde{v}_x) + h\, \|f' - \tilde{f}'\|_{L^p(]0,1[)}.$$

Indeed, by definition of $\mathcal{T}_h^f$ and $\mathcal{T}_h^{\tilde{f}}$, triangular inequality and convexity of W_p^p,

$$\begin{aligned} W_p([\mathcal{T}_h^f v]_x, [\mathcal{T}_h^{\tilde{f}} \tilde{v}]_x) &\le W_p(K_h * [T_h^f v]_x, K_h * [T_h^f \tilde{v}]_x) + W_p(K_h * [T_h^f \tilde{v}]_x, K_h * [T_h^{\tilde{f}} \tilde{v}]_x) \\ &\le W_p([T_h^f v]_x, [T_h^f \tilde{v}]_x) + W_p([T_h^f \tilde{v}]_x, [T_h^{\tilde{f}} \tilde{v}]_x), \end{aligned}$$

where the first term is bounded by $W_p(v_x, \tilde v_x)$ by Proposition 3.17, and the second term by $h\,\|f' - \tilde f'\|_{L^p(]0,1[)}$ by the first step.

3. Let v and $\tilde v$ be given in $\mathcal{U}$, and $t \geq 0$. Then

$$W_p([\mathcal{S}_h^f v]_x(t,.), [\mathcal{S}_h^{\tilde f} \tilde v]_x(t,.)) \leq W_p(v_x, \tilde v_x) + t\,\|f' - \tilde f'\|_{L^p(]0,1[)}.$$

Let us indeed decompose t as $(M+m)\,h$ with $M \in \mathbb{N}$ and $0 \leq m < 1$. Then, by induction from the second step,

$$W_p([(\mathcal{T}_h^f)^k v]_x, [(\mathcal{T}_h^{\tilde f})^k \tilde v]_x) \leq W_p(v_x, \tilde v_x) + k\,h\,\|f' - \tilde f'\|_{L^p(]0,1[)}$$

for any k in $\mathbb{N}$, and in particular for $k = M$ and $k = M+1$. Consequently, by definition of $\mathcal{S}_h$ and convexity, $W_p([\mathcal{S}_h^f v]_x(t,.), [\mathcal{S}_h^{\tilde f} \tilde v]_x(t,.))$ is bounded by

$$\begin{aligned}(1-m)\,W_p([(\mathcal{T}_h^f)^M v]_x, [(\mathcal{T}_h^{\tilde f})^M \tilde v]_x) + m\,W_p([(\mathcal{T}_h^f)^{M+1} v]_x, [(\mathcal{T}_h^{\tilde f})^{M+1} \tilde v]_x)&\\ \leq W_p(v_x, \tilde v_x) + t\,\|f' - \tilde f'\|_{L^p(]0,1[)}.&\end{aligned}$$

4. Proceeding as in Sections 3.5 and 3.6 we note that $[\mathcal{S}_h^f u^0]_x(t,.)$ and $[\mathcal{S}_h^{\tilde f} \tilde u^0]_x(t,.)$ respectively converge to $u_x(t,.)$ and $\tilde u(t,.)$ in the weak sense of probability measures in view of Propositions 3.21 when $\nu = 0$ or 3.33 when $\nu = 0$. In particular it follows from Proposition 3.15 and the third step of the current proof (with u^0 and $\tilde u^0$) that

$$W_p(u_x(t,.), \tilde u_x(t,.)) \leq \liminf_{h \to 0} W_p([\mathcal{S}_h^f v]_x(t,.), [\mathcal{S}_h^{\tilde f} \tilde v]_x(t,.)) \leq W_p(u_x^0, \tilde u_x^0) + t\,\|f' - \tilde f'\|_{L^p(]0,1[)}$$

for any $t \geq 0$. This concludes the argument. □

Deuxième partie

Inégalités de concentration et de transport

Chapitre 4

Inégalités de Csiszár-Kullback-Pinsker à poids

Ce chapitre correspond en grande partie à l'article [23] écrit en collaboration avec Cédric Villani et publié aux Annales de la Faculté des Sciences de Toulouse. Un complément est apporté en annexe.

Nous généralisons l'inégalité de Csiszár-Kullback-Pinsker en introduisant des fonctions de poids dans la variation totale, les poids admissibles dépendant de la décroissance à l'infini de la mesure de référence. A partir de cette nouvelle inégalité nous retrouvons en particulier l'équivalence d'une inégalité T_1 *et de l'existence d'un moment carré-exponentiel, dont nous donnons une autre démonstration, plus directe, en annexe. En application de ces résultats, nous établissons une variante des résultats de H. Djellout, A. Guillin et L. Wu [46] sur des inégalités de transport pour des systèmes dynamiques aléatoires, sous une condition portant sur des moments exponentiels. Comme autre conséquence nous retrouvons et généralisons un résultat de G. Blower [14] relatif à une perturbation d'une inégalité* T_2.

Introduction

Let X be an abstract Polish space, and let $P(X)$ be the set of all Borel probability measures on X ; let d be a lower semi-continuous metric on X, and let p belong to $[1, +\infty)$. Whenever μ, ν belong to $P(X)$, we define

• the **Wasserstein distance of order** p between μ and ν by

$$W_p(\mu, \nu) = \inf \left(\iint d(x, y)^p d\pi(x, y) \right)^{1/p}$$

where π runs over the set of probability measures on $X \times X$ with marginals μ and ν ;

• the **Kullback information** of μ with respect to ν by

$$H(\mu|\nu) = \int f \,\log f \,d\nu, \qquad f = \frac{d\mu}{d\nu};$$

by convention $H(\mu|\nu) = +\infty$ if μ is not absolutely continuous with respect to ν.

Both objects play an important role in a number of problems in probability theory, where they may be encountered under the names of Monge-Kantorovich distances, or minimal distances, and relative entropy, or relative H functional. More information can be found, together with many references, in [111]. For various purposes it is of interest to investigate whether they can be compared to each other. The most famous such inequality is the **Csiszár-Kullback-Pinsker inequality**, which we shall denote CKP inequality for short : if d is the trivial distance, i.e. $d(x, y) = 1_{x\neq y}$, then

$$2W_1(\mu, \nu) = \|\mu - \nu\|_{TV} \leq \sqrt{2H(\mu|\nu)},$$

where "TV" stands for the total variation norm.

Another class of inequalities which has been studied at length is encountered under the names of **Talagrand inequalities**, **transportation inequalities**, or **transportation cost-information inequalities** ; we shall just denote it by T_p. By definition, a reference probability measure ν satisfies the $T_p(\lambda)$ inequality for some $\lambda > 0$ if

$$\forall \mu \in P(X), \quad W_p(\mu, \nu) \leq \sqrt{\frac{2H(\mu|\nu)}{\lambda}};$$

and it satisfies T_p if it satisfies $T_p(\lambda)$ for some $\lambda > 0$. In particular, CKP inequality means that *any* reference probability measure satisfies $T_1(4)$ when d is the trivial distance.

We note right away that $W_p \leq W_{p'}$ for $p \leq p'$, so that T_p inequalities become stronger and stronger as p becomes larger. The cases $p = 1$ and $p = 2$ are of particular interest.

The study of T_p inequalities is a rather old topic [93], which recently received a new impulse. First, it was pointed out by Marton [77] and Talagrand [106] that these inequalities are a handy tool in the study of *concentration of measure* [68] ; in particular, Talagrand showed how to take advantage of the good tensorization properties of inequality T_2, to establish concentration in product spaces. At the same time, he established the validity of T_2 for the Gaussian measure, which justifies the terminology of "Talagrand inequalities". On the other hand, recent developments of the theory of optimal transportation led to new connections between these inequalities and other classes of functional inequalities with a geometric content, in particular **logarithmic Sobolev inequalities**. For instance, the main result in [88] is that *a logarithmic Sobolev inequality implies a T_2 inequality* (and the converse is also true under some convexity assumption). Various proofs and variants of these results, together with a detailed discussion, can be found in [16, 88, 111].

On the other hand, the works by Bobkov and Götze [17], and Djellout, Guillin and Wu [46] suggest that there is still room for investigation in an abstract Polish space setting, without any underlying geometric structure. More precisely, given a reference probability measure ν, one of the main results proven in these references is the equivalence between

1. ν satisfies a T_1 inequality ;
2. there exists λ such that $\displaystyle\int e^{t(f(x)-\int f(x)\,d\nu(x))}\,d\nu(x) \leq e^{\frac{t^2}{2\lambda}}$ for any real t and Lipschitz function f with Lipschitz seminorm 1 ;
3. ν admits a square-exponential moment, i.e. $\displaystyle\int e^{\alpha d(x,y)^2}\,d\nu(x)$ is finite for some $\alpha > 0$ and some (and thus any) y.

Notice how tractable is this criterium for T_1 : for instance, the validity of a logarithmic Sobolev inequality depends on subtle properties of the reference measure, which imply not only the existence of a square-exponential moment, but also – among other features which are still poorly identified – strict positivity, in a quantitative way which has not been made precise so far (see however [39] for important progress in that direction). Djellout, Guillin and Wu explored various applications of their result, including T_1 inequalities in path space for solutions of stochastic differential equations, or T_1 inequalities in large dimension for random dynamical systems under adequate assumptions of weak dependence.

The purpose of this paper is twofold.

On one hand, we shall establish a generalization of the CKP inequality, allowing for a weight in the total variation. How much weight is allowed will depend on the decay of the reference measure. In that generalization, the optimal constant 4 will be lost, but this will be more than compensated by the gain of precision brought by the weight. In view of the large range of applications of the usual CKP inequality, we do hope that this generalization can be of interest in various contexts.

On the other hand, we shall point out that, instead of considering CKP inequality as just a particular case of T_1, it is possible to establish many general comparison results between W_p and H by studying the weighted CKP inequality. In particular, we shall recover in a straightforward way (and with improved constants) the above-mentionned result according to which a square-exponential moment implies T_1. Then we shall establish a variant of the result by Djellout, Wu and Guillin [46] about random dynamical systems, in which assumptions are only expressed in terms of exponential moments. Not only are these conditions easier to check, but they also allow for more generality. Also, we shall establish weakened versions of T_1 and T_2 inequalities, in which the square-root on the right-hand side is replaced by a combination of powers, and which are satisfied with quite a bit of generality, under just decay assumptions on the reference measure. Among them is a generalization of an unpublished partial result by Blower [14] about the perturbation of T_2 inequalities.

The plan of the paper is as follows. In section 4.1, we state our weighted CKP inequality and derive from it various applications to the study of T_p inequalities and their variants. In section 4.2, we give a detailed proof of the weighted CKP inequality and in section 4.3 we show how our results can be applied to the study of discrete-time processes. Finally, in the appendix, we give another proof of the equivalence between a T_1 inequality and the existence of a square-exponential moment.

4.1 Main results

Working in a Polish space is a natural assumption when handling Wasserstein distances, because it is sufficient to derive all the well-known and useful properties of these distances, in particular their relation with the weak topology [111]. However, for all the results in this section, no use will be made of completeness or separability, and so we state the results with more generality.

In the sequel, the notation $\varphi(\mu - \nu)$ is a shorthand for the signed measure $\varphi\mu - \varphi\nu$.

Theorem 4.1 (weighted CKP inequalities). *Let X be a measurable space, let μ, ν be two probability measures on X, and let φ be a nonnegative measurable function on X. Then*

$$\text{(i)} \qquad \|\varphi\,(\mu-\nu)\|_{TV} \;\le\; \left(\frac{3}{2} + \log \int e^{2\varphi(x)}\,d\nu(x)\right)\left(\sqrt{H(\mu|\nu)} + \frac{1}{2}H(\mu|\nu)\right);$$

$$\text{(ii)} \qquad \|\varphi\,(\mu-\nu)\|_{TV} \;\le\; \sqrt{2}\left(1 + \log \int e^{\varphi(x)^2}\,d\nu(x)\right)^{1/2}\sqrt{H(\mu|\nu)}.$$

Remarks 4.2. 1. The assumption $\displaystyle\int_X e^{\varphi^2}\,d\nu < +\infty$ is always stronger than the assumption $\int_X e^{2\varphi}\,d\nu < +\infty$, so the inequality (i) above always applies in more generality than (ii). Further note that if we choose $\varphi \equiv 1$ in (ii), we recover the usual CKP inequality

$$\|\mu-\nu\|_{TV} \;\le\; c\sqrt{H(\mu|\nu)}$$

with the non-optimal constant $c = 2$ instead of $\sqrt{2}$. This shows that the constants on the right-hand side of (ii) cannot be improved by more than a factor $\sqrt{2}$. Although we worked quite a bit to decrease this numerical constant, it is likely that one can still do better, at least by replacing $\int e^{\varphi^2}$ with $\int e^{\lambda\varphi^2}$. Note though that the optimal constant $\sqrt{2}$ can be recovered by writing our proof again in the particular case $\varphi \equiv 1$, as it shall be pointed out in section 4.2.

2. Let us discuss very briefly the sharpness of the orders of magnitude in the above inequalities. When μ is very close to ν, the Kullback information can be approximated by a weighted squared L^2 norm, which shows that it is natural to expect a term in $\sqrt{H(\mu|\nu)}$ (as opposed to another power of H) in the right-hand side. On the other hand, consider the situation when $X = \mathbb{R}^n$, and the reference measure ν is the standard Gaussian distribution; choose $\varphi(x) = \delta|x|$ for $\delta < 1/\sqrt{2}$. Then the left-hand side of inequality (ii) will be typically $O(\sqrt{n})$ as $n \to \infty$, while the right-hand side will be typically $O(n)$. If $\varphi(x) = \delta\sum|x_i|/\sqrt{n}$, then the left-hand side will be typically $O(n)$, while the right-hand side will be typically $O(n^{3/2})$. These examples suggest that Theorem 4.1 still leaves room for improvement for problems set in large dimension. As we shall see in Section 4.3, this loss of a $O(\sqrt{n})$ factor will put limitation on the validity of measure concentration inequalities that can be deduced from Theorem 4.1 in large dimension.

We postpone the proof of Theorem 4.1 to the next section, and now list two consequences.

Corollary 4.3. *Let X be a measurable space equipped with a measurable distance d, let $p \geq 1$ and let ν be a probability measure on X. Assume that there exist $x_0 \in X$ and $\alpha > 0$ such that $\int e^{\alpha d(x_0,x)^p}\, d\nu(x)$ is finite. Then*

$$\forall \mu \in P(X), \qquad W_p(\mu,\nu) \leq C\Big[H(\mu|\nu)^{\frac{1}{p}} + \Big(\frac{H(\mu|\nu)}{2}\Big)^{\frac{1}{2p}}\Big],$$

where

$$C := 2 \inf_{x_0 \in X,\, \alpha > 0} \left(\frac{1}{\alpha}\Big(\frac{3}{2} + \log \int e^{\alpha d(x_0,x)^p}\, d\nu(x)\Big)\right)^{\frac{1}{p}} < +\infty.$$

Corollary 4.4. *Let X be a measurable space equipped with a measurable distance d, let $p \geq 1$ and let ν be a probability measure on X. Assume that there exist $x_0 \in X$ and $\alpha > 0$ such that $\int e^{\alpha d(x_0,x)^{2p}}\, d\nu(x)$ is finite. Then*

$$\forall \mu \in P(X), \qquad W_p(\mu,\nu) \leq C\, H(\mu|\nu)^{\frac{1}{2p}},$$

where

$$C := 2 \inf_{x_0 \in X,\, \alpha > 0} \left(\frac{1}{2\alpha}\Big(1 + \log \int e^{\alpha d(x_0,x)^{2p}}\, d\nu(x)\Big)\right)^{\frac{1}{2p}} < +\infty.$$

Particular case 4.5. When X is bounded, a simpler bound holds :

$$\forall \mu \in P(X), \qquad W_p(\mu,\nu) \leq 2^{\frac{1}{2p}} \operatorname{diam}(X)\, H(\mu|\nu)^{\frac{1}{2p}},$$

where $\operatorname{diam}(X) := \sup\{d(x,y);\ x,y \in X\}$.

Since the proofs of these results are very similar, we only give the proof of Corollary 4.4.

Proof of Corollary 4.4. On one hand it is known [111, Proposition 7.10] that

$$W_p^p(\mu,\nu) \leq 2^{p-1} \|d(x_0,.)^p\, (\mu-\nu)\|_{TV};$$

on the other hand the second part of Theorem 4.1 yields

$$\Big\|\sqrt{\alpha}\, d(x_0,.)^p\, (\mu-\nu)\Big\|_{TV} \leq \sqrt{2}\left(1 + \log \int e^{\alpha d(x_0,x)^{2p}}\, d\nu(x)\right)^{1/2} \sqrt{H(\mu|\nu)}.$$

This conclude the argument. $\square$

We now focus on some particular cases of interest, namely for $p = 1$ and $p = 2$ under assumptions of exponential moments of order 1, 2 and 4.

Corollary 4.6. *Let X be a measurable space equipped with a measurable distance d, let ν be a reference probability measure on X, and let x_0 be any element of X. Then*

(i) If $\int_X e^{\alpha d(x_0,x)}\, d\nu(x) < +\infty$ for some $\alpha > 0$, then there is a constant C such that

$$\forall \mu \in P(X), \qquad W_1(\mu,\nu) \le C\left(H(\mu|\nu) + \sqrt{H(\mu|\nu)}\right).$$

(ii) If $\int_X e^{\alpha d(x_0,x)^2}\, d\nu(x) < +\infty$ for some $\alpha > 0$, then there is a constant C such that

$$\forall \mu \in P(X), \qquad W_1(\mu,\nu) \le C\sqrt{H(\mu|\nu)};$$
$$\forall \mu \in P(X), \qquad W_2(\mu,\nu) \le C\left[\sqrt{H(\mu|\nu)} + H(\mu|\nu)^{\frac{1}{4}}\right].$$

In particular, ν satisfies T_1.

(iii) If $\int_X e^{\alpha d(x_0,x)^4}\, d\nu(x) < +\infty$ for some $\alpha > 0$, then there is a constant C such that

$$\forall \mu \in P(X), \qquad W_2(\mu,\nu) \le C\, H(\mu|\nu)^{\frac{1}{4}}.$$

Remark 4.7. Part (ii) of this corollary contains the result that the existence of an exponential moment of order 2 implies a T_1 inequality; according to [17], the converse is true, so this criterion is optimal. To compare these various results in practical situations, it is good to keep in mind the following elementary lemma :

Lemma 4.8. *Let X be a measurable space equipped with a measurable distance d, let $p \ge 1$ and let ν be a probability measure on X. Then the following three statements are equivalent :*

1. *there exist $x_0 \in X$ and $\alpha > 0$ such that $\int e^{\alpha d(x_0,x)^p}\, d\nu(x)$ is finite;*
2. *for any $x_0 \in X$, there exists $\alpha > 0$ such that $\int e^{\alpha d(x_0,x)^p}\, d\nu(x)$ is finite;*
3. *there exists $\alpha > 0$ such that $\iint e^{\alpha d(x,y)^p}\, d\nu(x)\, d\nu(y)$ is finite.*

Moreover,

$$\inf_{x_0\in X} \int e^{\alpha d(x_0,x)^p}\, d\nu(x) \le \iint e^{\alpha d(x,y)^p}\, d\nu(x)\, d\nu(y) \le \left(\inf_{x_0\in X}\int e^{\alpha 2^{p-1} d(x_0,x)^p}\, d\nu(x)\right)^2.$$

Remark 4.9. The following two results can be deduced from the equivalence between the existence of an exponential moment of order 2 and a T_1 inequality :

1. Let μ be a probability measure on a Polish space X, satisfying T_1. Then so does any probability measure $\nu = h\mu$, where h is a μ-almost surely bounded measurable function on X.

2. Let μ be a probability measure on $\mathbb{R}^d$ satisfying T_1. Then so does its marginal (via orthogonal projection) on any hyperplane of $\mathbb{R}^d$.

Remark 4.10. Part (ii) also generalizes the perturbation result proven by Blower, who showed in [14] that an inequality of the form $W_2 \leq C(H^{1/2} + H^{1/4})$ holds true when ν is bounded from above and below by constant multiples of a reference measure ν_0 satisfying T_2. In fact, if ν_0 satisfies T_2, then it also satisfies T_1, so it has a finite square-exponential moment, and so does ν if it is bounded above by a constant multiple of ν_0.

Remark 4.11. Let ν be a reference probability measure having finite exponential moments of order p; how far is it from satisfying T_p? The preceding results indicate that the answer is very different for $p = 1$ and $p = 2$. If T_1 is not satisfied, this means that the decay of ν at infinity is not fast enough, and the T_1 inequality usually fails for *large* values of the Kullback information. On the contrary, if T_2 is not satisfied, this is not necessarily just for a question of fast decay (remember that T_2 implies strict positivity), and the T_2 inequality usually fails for *small* values of the Kullback information. In particular, it is no wonder that we did not manage to recover T_2 inequalities with our arguments taking into account only the decay of ν.

4.2 Proof of the main inequalities

We shall now present detailed proofs of the main inequalities in Theorem 1.

Proof of Theorem 4.1.

Without loss of generality, we assume that μ is absolutely continuous with respect to ν, with density f. We set $u := f - 1$, so that

$$\mu \;=\; (1+u)\nu;$$

we note that $u \geq -1$ and $\displaystyle\int_X u\, d\nu \;=\; 0$. We also define

$$h(v) \;:=\; (1+v)\log(1+v) - v, \qquad v \in [-1, +\infty)$$

so that

$$H(\mu|\nu) \;=\; \int_X h(u)\, d\nu. \tag{4.1}$$

We note that $h \geq 0$.

We start with the **proof of inequality (i)**, splitting the weighted total variation as

$$\int \varphi\, d|\mu - \nu| \;=\; \int \varphi\, |u|\, d\nu \;=\; \int_{\{-1 \leq u \leq 4\}} \varphi\, |u|\, d\nu \;+\; \int_{\{u > 4\}} \varphi\, u\, d\nu. \tag{4.2}$$

We shall estimate both terms separately, first bounding the **first term** ($u \leq 4$) in (4.2). By Cauchy-Schwarz inequality,

$$\int_{u \leq 4} \varphi\, |u|\, d\nu \leq \left(\int_{u \leq 4} \varphi^2\, d\nu\right)^{1/2} \left(\int_{u \leq 4} u^2\, d\nu\right)^{1/2}.$$

On the other hand, from the elementary inequality

$$-1 \le v \le 4 \quad \Longrightarrow \quad v^2 \le 4h(v)$$

(a consequence of the fact that $h(v)/v$ is nondecreasing) we deduce

$$\int_{u\le 4} u^2\,d\nu \;\le\; 4\int_{u\le 4} h(u)\,d\nu.$$

Combining this with the nonnegativity of h and (4.1), we find

$$\int_{u\le 4}\varphi\,|u|\,d\nu \;\le\; 2\left(\int_X \varphi^2\,d\nu\right)^{1/2}\left(\int_X h(u)\,d\nu\right)^{1/2} = 2\left(\int_X \varphi^2\,d\nu\right)^{1/2} H(\mu|\nu)^{1/2}. \tag{4.3}$$

Since the function $t \longmapsto e^{2\sqrt{t}}$ is increasing and convex on $[1/4,+\infty)$ we can write

$$\exp\left(2\sqrt{\int_X \varphi^2\,d\nu}\right) \le \exp\left(2\sqrt{\int_X (\varphi+1/2)^2\,d\nu}\right) \le \int_X \exp\left(2\sqrt{(\varphi+1/2)^2}\right)\,d\nu$$
$$= \int_X e^{2\varphi+1}\,d\nu.$$

In other words,

$$2\sqrt{\int_X \varphi^2\,d\nu} \;\le\; 1 \,+\, \log\int e^{2\varphi}\,d\nu;$$

if we plug this into (4.3), we conclude that

$$\int_{u\le 4}\varphi\,|u|\,d\nu \;\le\; \left(1 \,+\, \log\int_X e^{2\varphi}\,d\nu\right)\, H(\mu|\nu)^{1/2}. \tag{4.4}$$

We now turn to the estimate of the **second term** ($u > 4$) in (4.2). By applying the Young-type inequality

$$w\xi \;\le\; w\log w - w + e^{\xi} \qquad (w\ge 0, \xi\in\mathbb{R}) \tag{4.5}$$

with $w = u(x)$ and $\xi = \varphi(x) - Z$, where Z is a nonnegative constant to be chosen later, we find

$$\begin{aligned} u(x)\varphi(x) \;&\le\; u(x)\log u(x) \,-\, u(x) \,+ e^{\varphi(x)-Z} \,+ Zu(x)\\ &\le h(u(x)) \,+\, \left(\inf_{v>4}\sqrt{h(v)}\right)^{-1} e^{\varphi(x)-Z}\sqrt{h(u(x))} \,+ Zu(x)\end{aligned}$$

on $\{u(x) > 4\}$. By integration, we deduce

$$\int_{u>4} u\varphi\,d\nu \;\le\; \int_{u>4} h(u)\,d\nu + \sqrt{k}\int_{u>4} e^{\varphi-Z}\sqrt{h(u)}\,d\nu + Z\int_{u>4} u\,d\nu,$$

where

$$k := \Big(\inf_{v>4} h(v)\Big)^{-1} = \frac{1}{h(4)} < \frac{1}{4}.$$

By Cauchy-Schwarz inequality again,

$$\int_{u>4} e^{\varphi-Z}\,\sqrt{h(u)}\,d\nu \;\le\; \sqrt{\int_X e^{2(\varphi-Z)}\,d\nu}\,\sqrt{\int_{u>4} h(u)\,d\nu} = \sqrt{\int_X e^{2(\varphi-Z)}\,d\nu}\,\sqrt{H(\mu|\nu)}.$$

Finally, from the inequality

$$v \ge 4 \quad\Longrightarrow\quad v \le 4k\,h(v)$$

we deduce

$$\int_{u>4} u\,d\nu \;\le\; 4k\int_{u>4} h(u)\,d\nu \le 4kH(\mu|\nu).$$

Our conclusion is that, for any constant $Z \ge 0$,

$$\int_{u>4} \varphi\,u\,d\nu \;\le\; (1+4kZ)H(\mu|\nu) \,+\, \sqrt{k}\,\sqrt{\int_X e^{2(\varphi-Z)}\,d\nu}\,\sqrt{H(\mu|\nu)}. \tag{4.6}$$

We now choose Z in such a way that

$$\int_X e^{2(\varphi-Z)}\,d\nu \;=\; 1;$$

in other words,

$$Z \;:=\; \frac{1}{2}\log\int e^{2\varphi}\,d\nu \ge 0.$$

Plugging this into (4.6), we conclude that

$$\int_{u>4} \varphi\,u\,d\nu \;\le\; \left(1+2k\log\int e^{2\varphi}\,d\nu\right) H(\mu|\nu) + \,\sqrt{k}\,\sqrt{H(\mu|\nu)}. \tag{4.7}$$

Now inequality (i) follows from (4.4) and (4.7) upon noting that $1+\sqrt{k} < \dfrac{3}{2}$ and $2k < \dfrac{1}{2}\cdot$

We next turn to the **proof of (ii)**. Although the decomposition (4.2) and the same kind of argument would also lead to the result, we prefer to proceed as follows.

Since $h(0) = h'(0) = 0$, by Taylor's formula with integral remainder, we can write

$$h(u) \;=\; u^2\int_0^1 \frac{1-t}{1+tu}\,dt,$$

and thus

$$H(\mu|\nu) \;=\; \int_X\int_0^1 \frac{u^2(x)\,(1-t)}{1+tu(x)}\,d\nu(x)\,dt.$$

On the other hand, by Cauchy-Schwarz inequality on $(0,1)\times X$

$$\left(\int_0^1 (1-t)\,dt\right)^2 \left(\int_X \varphi|u|\,d\nu\right)^2 = \left(\int_X\int_0^1 (1-t)\varphi|u|\,d\nu\,dt\right)^2$$

$$\leq \left[\int_X\int_0^1 (1-t)\,(1+tu)\,\varphi^2\,d\nu\,dt\right]\left[\int_x\int_0^1 \frac{1-t}{1+tu}|u|^2\,d\nu\,dt\right];$$

thus

$$\left(\int_X \varphi\,|u|\,d\nu\right)^2 \leq CH(\mu|\nu)$$

where

$$C := \frac{\displaystyle\iint (1-t)\,(1+tu)\,\varphi^2\,d\nu\,dt}{\left(\displaystyle\int_0^1 (1-t)\,dt\right)^2}\cdot \tag{4.8}$$

We decompose the numerator as follows :

$$\iint (1-t)\,(1+tu)\,\varphi^2\,d\nu\,dt = \int (1-t)t\,dt\int (1+u)\,\varphi^2\,d\nu + \int (1-t)^2\,dt\int \varphi^2\,d\nu$$

$$= \frac{1}{6}\int \varphi^2\,d\mu + \frac{1}{3}\int \varphi^2\,d\nu. \tag{4.9}$$

From the convexity inequality

$$\int \varphi^2\,d\mu \leq H(\mu|\nu) + \log\int e^{\varphi^2}\,d\nu, \tag{4.10}$$

(a well-known consequence of (4.5), see for instance [68, eq. (5.13)]) and Jensen's inequality, in the form

$$\int \varphi^2\,d\nu \leq \log\int e^{\varphi^2}\,d\nu, \tag{4.11}$$

we deduce that the right-hand side of (4.9) is bounded above by

$$\frac{1}{6}H(\mu|\nu) + \frac{1}{2}\log\int e^{\varphi^2}\,d\nu.$$

Plugging this into (4.8), we conclude that

$$\left(\int \varphi\,|u|\,d\nu\right)^2 \leq \left(\frac{2}{3}H + 2L\right)H, \tag{4.12}$$

where H stands for $H(\mu|\nu)$ and L for $\log\displaystyle\int e^{\varphi^2}\,d\nu$.

The preceding bound is good only for "small" values of H. We now complement it with another bound which is relevant for "large" values of H. To do so, we write

$$\begin{aligned}\left(\int \varphi|u|\,d\nu\right)^2 &\le \int \varphi^2|u|\,d\nu \int |u|\,d\nu \\ &\le \left(\int \varphi^2\,d\mu + \int \varphi^2\,d\nu\right)\left(\int d\mu + \int d\nu\right) \\ &\le (H+2L)\,2\end{aligned}$$

where we have successively used Cauchy-Schwarz inequality, the inequality $|u| \le 1+u+1$ on $[-1,+\infty)$ (which results in $|u|\,\nu \le \mu+\nu$), and finally (4.10) and (4.11).

Combining this with (4.12), we obtain

$$\left(\int \varphi|u|\,d\nu\right)^2 \le \min\left((2H)\big(\frac{H}{3}+L\big)\,,\,2(H+2L)\right).$$

From the elementary inequality

$$\min(at^2+bt,\,t+d) \,\le\, Mt, \qquad M \,=\, \frac{1}{2}\left\{1+b+\sqrt{(b-1)^2+4ad}\right\}$$

we get

$$\int \varphi|u|\,d\nu \,\le\, m\sqrt{H(\mu|\nu)}$$

where

$$m \le \sqrt{1+L+\sqrt{(L-1)^2+\frac{8}{3}L}} \,\le \sqrt{2}\,\sqrt{L+1}.$$

This concludes the proof. □

Remark 4.12. If $\varphi \equiv 1$, we can replace the inequality (4.10) by just $\int_x d\mu = 1$; then the first part of the proof of (ii) becomes a proof of the usual CKP inequality, with the sharp constant $\sqrt{2}$.

4.3 Application to random dynamical systems

Let now be given a Polish space X, an arbitrary element $x_0 \in X$ and a set of conditional Borel probability measures $(P_k(\,\cdot\,|x^{k-1}))_{x^{k-1}\in X^{k-1},\,k\ge 1}$, depending on $x^{k-1} = (x_1,\ldots,x_{k-1}) \in X^{k-1}$ in a measurable way. We interpret x_0 as the (deterministic) initial position of a random dynamical system $(X_k)_{k\in\mathbb{N}}$, with values in X, and $P_k(\,\cdot\,|x^{k-1})$ as the law of X_k, knowing that $X_0 = x_0$ and $(X_1,\ldots,X_{k-1}) = x^{k-1}$. The question is whether it is possible, knowing some nice bounds on the conditional probability measures, to get a T_1 inequality for the law P^n of $(X_1,\ldots,X_n)$ on X^n, with a nice dependence on n.

Let us first assume that all the conditional probability measures satisfy a T_1 inequality, say with a uniform constant. In the context of independent random variables, it is rather easy [68, p. 122] to show that P^n satisfies $T_1(\lambda)$ for $\lambda^{-1} = O(n)$, and that this is sharp in general. Now we want to know whether the same behavior is generic for dependent random variables. Some results in that direction have been obtained by Marton and by Rio ; they are summarized and slightly improved in [46]. In those references it is shown that if each $P_k(\,\cdot\,|x^{k-1})$ satisfies $T_1(\kappa)$ for some fixed $\kappa > 0$, and the random dynamical system is weakly dependent, in the sense that the future does not depend too much on the present, then the answer is positive. See [46, Section 4] for precise assumptions. For instance, a sufficient condition is that the dynamical system is Markovian and that the map

$$x_{k-1} \longmapsto P_k(\ \cdot\ |x^{k-1})$$

is L-Lipschitz from X to $P(X)$, equipped with the W_1 distance, uniformly in k, for some $L < 1$.

In the present section, we shall establish a variant of this result under a different set of assumptions, which seems to be easier to check in practical situations, because it is expressed in terms of exponential moments with respect to a given origin point (which we chose, arbitrarily, as the starting point of the dynamical system). What will make our argument work (in a very straightforward way) is the simple and explicit dependence of the constants in Theorem 4.1 upon n when X is replaced by X^n.

In the sequel we consider a Polish space X, equipped with a measurable distance d, x_0 an arbitrary element in X, and $(P_k(\,\cdot\,|x^{k-1}))_{x^{k-1}\in X^{k-1},\,k\geq 1}$ a family of Borel probability measures on X, depending on $x^{k-1} := (x_1, \ldots, x_{k-1}) \in X^{k-1}$ in a measurable way. For all $n \geq 1$, we define the probability measure P^n on X^n by

$$dP^n(x_1, \ldots, x_n) = dP_1(x_1)\, dP_2(x_2|x_1)\, \ldots\, dP_n(x_n|x_1, \ldots, x_{n-1}),$$

and equip X^n with the distance D defined by

$$D(x, y) = D_2(x, y) := \sqrt{\sum_{k=1}^{n} d(x_k, y_k)^2}.$$

There is an important difference with the above-mentioned works, namely the choice of the distance on the product space X^n : instead of D_2, they consider the distance

$$D_1(x, y) := \sum_{k=1}^{n} d(x_k, y_k).$$

While D_2 is often more natural than D_1, the latter is better adapted for arguments involving tensorization and Lipschitz functions. Of course, $D_1 \leq \sqrt{n} D_2$, so the distance D_2 is stronger than D_1 for each finite n, but does not behave similarly in the asymptotic regime $n \to +\infty$. Accordingly, if we try to deduce natural concentration estimates from our results, we typically obtain

$$P^n\left[\left|\frac{1}{n}\sum_{k=1}^{n}\varphi(x_k) - \int\left(\frac{1}{n}\sum_{k=1}^{n}\varphi(x_k)\right)\, dP^n(x^n)\right| \geq \varepsilon\right] \leq 2\exp\left(-\frac{\lambda\varepsilon^2}{2}\right)$$

for any $n \geq 1$ and any Lipschitz function φ on X with Lipschitz seminorm 1. The fact that this bound does not go to 0 as $n \to \infty$ is probably linked to Remark 4.2 (2).

Theorem 4.13 (T_1 inequalities for random dynamical systems). *With the above notation, assume the existence of $\alpha_0 > 0$, a sequence $(z_k)_{k\geq 1}$ in X and families of nonnegative numbers $(\gamma_k)_{k\geq 1}$, $(\beta_j)_{j\geq 1}$ with*

$$\gamma := \sup_{n\geq 1}\left[\frac{1}{n}\sum_{k=1}^{n}\gamma_k\right] < +\infty, \qquad \beta := \sum_{j\geq 1}\beta_j < \alpha_0,$$

such that for all $k \geq 1$, $x^{k-1} \in X^{k-1}$,

$$\log \int_X e^{\alpha_0\, d(z_k,x_k)^2}\, dP_k(x_k|x^{k-1}) \leq \gamma_k + \sum_{j=1}^{k-1}\beta_j\, d(z_{k-j}, x_{k-j})^2.$$

Then, there exists $\lambda > 0$ such that for all $n \geq 1$, P^n satisfies $T_1(\lambda/n)$.

Particular case 4.14. Consider a homogeneous Markov chain on X with transition kernel $P(dy|x)$. Assume the existence of $(x_0, y_0) \in X \times X$, $\alpha_0 > 0$, $\beta < \alpha_0$ and $C < +\infty$ such that

$$\forall x \in X, \qquad \int_X e^{\alpha_0\, d(y_0,y)^2}\, P(dy|x) \leq Ce^{\beta\, d(x_0,x)^2}. \tag{4.13}$$

Then there exists $\lambda > 0$ such that for all $n \geq 1$, P^n satisfies $T_1(\lambda/n)$.

Remark 4.15. If Condition (4.13) is satisfied for some choice of $(x_0, y_0, \alpha_0, \beta, C)$, then for any $\alpha_0' < \alpha_0$ and $(x_0', y_0') \in X \times X$ we can find $\beta' \in [\beta, \alpha_0')$, $C' < +\infty$ such that Condition (4.13) is satisfied for $(x_0', y_0', \alpha_0', \beta', C')$. Thus the choice of reference points x_0 and y_0 is arbitrary : for instance, if $X = \mathbb{R}^d$, we can choose 0 for both, and the condition becomes

$$\exists \alpha > 0,\ \beta < \alpha,\ C < +\infty; \qquad \forall x \in \mathbb{R}^d, \qquad \int_{\mathbb{R}^d} e^{\alpha\,|y|^2}\, P(dy|x) \leq Ce^{\beta\,|x|^2}. \tag{4.14}$$

Proof of Theorem 4.13. Let $\alpha := \alpha_0 - \beta$. Since $\alpha \leq \alpha_0$, by assumption,

$$\int_X e^{\alpha d(x_0,x_n)^2}\, P_n(dx_n|x^{n-1}) \leq \exp\left(\gamma_n + \sum_{k=1}^{n-1}\beta_k d(z_{n-k}, x_{n-k})^2\right).$$

In particular,

$$\int_{X^n} e^{\alpha D(z^n,x^n)^2}\, P^n(dx^n) \leq e^{\gamma_n}\int_{X^n} \exp\left(\sum_{k=1}^{n-1}(\alpha+\beta_k)d(z_{n-k}, x_{n-k})^2\right)\, P^{n-1}(dx^{n-1}).$$

Here $z^n = (z_1, \dots, z_n)$; note that $\alpha + \beta_k \le \alpha_0$ for all k, and in particular we can repeat the argument with $n-1$ in place of n. Using an induction argument, one easily shows that

$$\int_{X^n} e^{\alpha D(x^n,z^n)^2}\, P^n(dx^n) \le e^{\sum_{k=1}^n \gamma_k} \le e^{n\gamma}.$$

In particular,

$$\log \int_{X^n} e^{\alpha D(x^n,z^n)^2}\, P^n(dx^n) = O(n),$$

and we conclude by applying the results presented in section 4.1. □

As examples of application we now consider the following two particular cases :

Example 4.16. Let (X_i) be a Markovian dynamical system on a Polish space X, with transition kernel $P(\,\cdot\,|x)$ such that

(i) $P(\,\cdot\,|x)$ satisfies $T_1(\lambda)$ for a constant λ independent of x ;

(ii) the map $x \mapsto P(\cdot\,|x)$ is L-Lipschitz from X to $P(X)$, equipped with the W_1 distance, with $L < 1$.

Then there exist $\alpha > 0$ and $\beta < \alpha$ such that for any $x_0, y_0 \in X$, there exists $\gamma < +\infty$ such that

$$\log \int_X e^{\alpha\, d(y_0,y)^2}\, P(dy|x) \le \gamma + \beta\, d(x_0,x)^2$$

for all $x \in X$. In particular the hypotheses of Theorem 4.13 hold in view of the Particular case 4.14.

Example 4.17. Let $(X_k)_{k\in\mathbb{N}}$ be a dynamical system on $\mathbb{R}^d$ such that the hypotheses of [46, Theorem 4.1] hold, that is, with the notation introduced above,

(i) there exists some constant λ such that

$$W_1(\nu, P_k(\,\cdot\,|x^{k-1})) \le \sqrt{\frac{2}{\lambda} H\big(\nu|P_k(\,\cdot\,|x^{k-1})\big)}$$

for all $k \ge 1$, x^{k-1} in $(\mathbb{R}^d)^{k-1}$ and all probability measures ν on $\mathbb{R}^d$;

(ii) there exist some nonnegative numbers a_j such that $\displaystyle\sum_{j=1}^{+\infty} a_j < 1$ and

$$W_1\big(P_k(\,\cdot\,|x^{k-1}), P_k(\,\cdot\,|\tilde{x}^{k-1})\big) < \sum_{j=1}^{k-1} a_j |x_{k-j} - \tilde{x}_{k-j}|$$

for all $k \ge 1$ and $x^{k-1}, \tilde{x}^{k-1}$ in $(\mathbb{R}^d)^{k-1}$.

Then the assumptions of Theorem 4.13 also hold for this system.

This last example shows that our assumptions are not less general than those in [46]. Note carefully that when we apply Theorem 4.13 to this system, we do not recover such a strong conclusion as in [46] because of the choice of distances on product spaces (D_2 instead of D_1).

Since the proofs for both Examples 4.16 and 4.17 are similar, we only study the second example.

Proof of the assertion in Example 4.17. In a **first step** we prove that for any $k \geq 1$, x^{k-1}, z^{k-1} in $(\mathbb{R}^d)^{k-1}$, z_k in $\mathbb{R}^d$, $\varepsilon, \delta > 0$ and $a < \dfrac{\lambda}{2}$, we have

$$\begin{aligned}\log \int & e^{a(1-\varepsilon)|y_k - z_k|^2}\, P_k(dy_k|z^{k-1}) \\ & \leq -\frac{1}{2}\log(1-\frac{2a}{\lambda}) + a\left(\frac{1}{\varepsilon}-1\right)\left(1+\frac{1}{\delta}\right)\left(\int |t_k - z_k|\, P_k(dt_k|z^{k-1})\right)^2 \\ & \qquad + a\left(\frac{1}{\varepsilon}-1\right)(1+\delta)\sum_{j=1}^{k-1} a_{k-j}\sum_{j=1}^{k-1} a_{k-j}|x_j - z_j|^2.\end{aligned}$$

Indeed, the probability measure $P_k(\,\cdot\,|x^{k-1})$ satisfies $T_1(\lambda)$ and the map $y \mapsto |y - z_k|$ is 1-Lipschitz, so by the Bobkov-Götze formulation of the T_1 inequality (see [17, Theorem 1.3] and [46, Section 1]) we have

$$\int e^{a\left[|y_k - z_k| - \int |t_k - z_k|\, P_k(dt_k|x^{k-1})\right]^2} P_k(dy_k|x^{k-1}) \;\leq\; \frac{1}{\sqrt{1-2a/\lambda}} \tag{4.15}$$

for any $a < \dfrac{\lambda}{2}$, $z_k \in \mathbb{R}^d$ and $x^{k-1} \in (\mathbb{R}^d)^{k-1}$.

Let then ε be some positive number. Integrating the inequality

$$(1-\varepsilon)|y_k - z_k|^2 \leq \left||y_k - z_k| - \int |t_k - z_k|\, P_k(dt_k|x^{k-1})\right|^2 + \left(\frac{1}{\varepsilon}-1\right)\left(\int |t_k - z_k|\, P_k(dt_k|x^{k-1})\right)^2$$

and using (4.15) lead to

$$\log \int e^{a(1-\varepsilon)|y_k - z_k|^2}\, P_k(dy_k|x^{k-1}) \;\leq\; -\frac{1}{2}\log(1-\frac{2a}{\lambda}) + a\left(\frac{1}{\varepsilon}-1\right)\left(\int |t_k - z_k|\, P_k(dt_k|x^{k-1})\right)^2.$$

Recall the Kantorovich-Rubinstein formulation of the W_1 distance [111, Theorem 1.14] :

$$W_1(\mu,\nu) = \sup_{g\ 1-\mathrm{Lipschitz}} \left(\int g\, d\mu - \int g\, d\nu\right)$$

This and Assumption (ii), with $\tilde{x}^{n-1} = z^{n-1}$, imply

$$\begin{aligned}\int |t_k - z_k|\, P_k(dt_k|x^{k-1}) - \int |t_k - z_k|\, P_k(dt_k|z^{k-1}) \;&\leq\; W_1(P_k(\,\cdot\,, x^{k-1}), P_k(\,\cdot\,, z^{k-1})) \\ &\leq\; \sum_{j=1}^{k-1} a_{k-j}\, |x_j - z_j|.\end{aligned}$$

Thus for any positive number δ

$$\left(\int |t_k - z_k|\, P_k(dt_k|x^{k-1})\right)^2 \le \left(1+\frac{1}{\delta}\right)\left(\int |t_k - z_k|\, P_k(dt_k|z^{k-1})\right)^2 + (1+\delta)\left(\sum_{j=1}^{k-1} a_{k-j}\,|x_j - z_j|\right)^2,$$

and by Cauchy-Schwarz inequality we can bound this quantity by

$$\left(1+\frac{1}{\delta}\right)\left(\int |t_k - z_k|\, P_k(dt_k|z^{k-1})\right)^2 + (1+\delta)\sum_{j=1}^{k-1} a_{k-j}\sum_{j=1}^{k-1} a_{k-j}\,|x_j - z_j|^2.$$

This concludes this first step.

In the **second step**, we build the sequence (z_k) by the following induction process. Let z_1 be arbitrary in $\mathbb{R}^d$; assuming that we have defined $z^{k-1} = (z_1, \dots, z_{k-1})$, we let

$$z_k := \int_{\mathbb{R}^d} t_k\, P_k(dt_k|z^{k-1}).$$

Then

$$\begin{aligned} a\left(\int |t_k - z_k|\, P_k(dt_k|z^{k-1})\right)^2 &\le \log \int e^{a\,|t_k - z_k|^2}\, P_k(dt_k|z^{k-1}) \\ &= \log \int e^{a\,\left|t_k - \int t_k\, P_k(dt_k|z^{k-1})\right|^2}\, P_k(dt_k|z^{k-1}) \\ &\le \log \frac{1}{\sqrt{1-2a/\lambda}} \end{aligned}$$

thanks to Jensen's inequality and again the Bobkov-Götze formulation of the T_1 inequality, which is satisfied by $P_k(\,\cdot\,|z^{k-1})$.

Now we choose $\varepsilon \in (0,1)$ and $\delta > 0$ in such a way that

$$a\left(\frac{1}{\varepsilon}-1\right)(1+\delta)\sum_{j=1}^{+\infty} a_j \;< a(1-\varepsilon):$$

for instance, $\varepsilon := \left(\sum_{j=1}^{+\infty} a_j\right)^{1/2}$ and $\delta := \left(\sum_{j=1}^{+\infty} a_j\right)^{-1/4} - 1$ will do. Then the assumptions of Theorem 4.13 can be checked to hold for

$$\alpha_0 := a(1-\varepsilon),$$

$$\gamma_k := -\frac{1}{2}\log\left(1-\frac{2a}{\lambda}\right)\left[1+\left(\frac{1}{\varepsilon}-1\right)\left(1+\frac{1}{\delta}\right)\right]$$

and

$$\beta_j := a\left(\frac{1}{\varepsilon}-1\right)(1+\frac{1}{\delta})\,a_j.$$

□

4.4 Appendix : A direct proof of the characterization of T_1 inequalities

According to what is above, a Borel probability measure ν on a Polish space (X,d) satisfies a T_1 inequality if and only if it admits a square-exponential moment. In particular, if there exist $\alpha > 0$ and $x_0 \in X$ such that $\int_X e^{\alpha d(x,x_0)^2}\, d\nu(x)$ be finite, then, by Corollary 4.4, ν satisfies the inequality

$$W_1(\mu,\nu) \le C\, H(\mu|\nu)^{\frac{1}{2}}$$

for all μ where

$$C := \inf_{x_0\in X,\ \alpha>0} \left(\frac{2}{\alpha}\Big(1 + \ln \int_X e^{\alpha d(x_0,x)^2}\, d\nu(x)\Big)\right)^{\frac{1}{2}}.$$

In other words ν satisfies $T_1(\lambda)$ with $\lambda = \sup_{x_0\in X,\ \alpha>0} \left(\frac{1}{\alpha}\Big(1 + \ln \int e^{\alpha d(x_0,x)^2}\, d\nu(x)\Big)\right)^{-1}$.

In this appendix we give a direct proof of this implication (with a slightly worse constant), based on an idea communicated to us by M. Ledoux.

Theorem 4.18. *Let ν be a Borel probability measure on a Polish space (X,d) such that $\int_X e^{\alpha d(x,x_0)^2}\, d\nu(x)$ be finite for some $\alpha > 0$ and $x_0 \in X$. Then ν satisfies the inequality*

$$W_1(\mu,\nu) \le C\, H(\mu|\nu)^{\frac{1}{2}}$$

for all μ where

$$C := 3.07 \inf_{x_0\in X,\ \alpha>0} \left(\frac{2}{\alpha} \sup\Big(0.7, \ln \int_X e^{\alpha d(x_0,x)^2}\, d\nu(x)\Big)\right)^{\frac{1}{2}}.$$

Proof. By the Bobkov-Götze formulation of the T_1 inequality it is sufficient to prove that

$$\int_X e^{tf(x)}\, d\nu(x) \;\le e^{\frac{C^2 t^2}{4}}$$

holds for any $t \in \mathbb{R}$ and any Lipschitz function f on X with Lipschitz seminorm 1 and zero mean $\int_X f(x)\, d\mu(x)$, where C is given as in the Theorem.

For this purpose we fix such a function f, and let

$$\Lambda(t) = \ln \int_X e^{tf(x)}\, d\nu(x), \qquad t \in \mathbb{R} :$$

in this notation we wish to prove that the inequality

$$\Lambda(t) \le \frac{C^2\, t^2}{4}$$

holds for any real number t, or at least any nonnegative t since we may change f into $-f$.

1. We let $\ell > 0$ to be fixed later on and choose $k > 0$ such that

$$\int_X e^{k^2 f^2(x)}\, d\nu(x) \le e^{2\ell}, \tag{4.16}$$

for instance $k^2 = \dfrac{\alpha}{2} \inf\left(\dfrac{\ell}{L}, 1\right)$ where $L = \ln \displaystyle\int_X e^{\alpha d(x_0,x)^2}\, d\nu(x)$.

Indeed, by triangular, Young's, Jensen's and Hölder's inequalities,

$$\begin{aligned}
\int_X e^{k^2 f^2(x)}\, d\nu(x) &= \int_X e^{k^2\left[\int (f(x)-f(y))\, d\nu(y)\right]^2} d\nu(x) \\
&\le \int_X e^{k^2\left[\int d(x,y) d\nu(y)\right]^2} d\nu(x) \\
&\le e^{2k^2 \int d(x_0,y)^2\, d\nu(y)} \int_X e^{2k^2 d(x,x_0)^2}\, d\nu(x) \\
&\le \left(\int_X e^{2k^2 d(x,x_0)^2}\, d\nu(x)\right)^2 \\
&\le \left(\int_X e^{\alpha d(x,x_0)^2}\, d\nu(x)\right)^{\frac{4k^2}{\alpha}} = e^{\frac{4Lk^2}{\alpha}} \le e^{2\ell}.
\end{aligned}$$

2. For any $p \ge 1$ and $t \ge 0$ we have

$$\Lambda(t) \le \frac{1}{p} \ln \int_X e^{ptf(x)}\, d\nu(x) \le \frac{1}{p} \ln \int_X e^{\frac{p^2t^2}{4k^2} + k^2 f^2(x)} d\nu(x)$$

by Hölder's and Young's inequalities, so that

$$\Lambda(t) \le \frac{p\, t^2}{4\, k^2} + \frac{2\,\ell}{p}\cdot \tag{4.17}$$

3. By assumption on f and ν, Taylor's formula ensures that

$$\Lambda(t) \le \frac{t^2}{2} \sup_{s\in[0,t]} \Lambda''(s)$$

where

$$\Lambda''(t) \le \int_X f^2(x)\, e^{tf(x)}\, d\nu(x)$$

since

$$\int_X e^{tf(x)}\, d\nu(x) \ge \exp\Big(t \int_X f(x)\, d\nu(x)\Big) = 1$$

by Jensen's inequality and assumption on f.

But, for any $\varepsilon > 0$,

$$e^{tf(x)} \le e^{\frac{\varepsilon t^2}{2k^2}}\, e^{\frac{k^2 f^2}{2\varepsilon}}$$

by Young's inequality and, for any $\varepsilon > 1/2$,

$$f^2 \leq \frac{2\varepsilon}{2\,\varepsilon - 1}\,\frac{1}{e\,k^2}\,e^{\left(1-\frac{1}{2\varepsilon}\right)k^2 f^2}$$

by the inequality $e\,x \leq e^x$. Hence

$$\Lambda(t) \leq \frac{\varepsilon}{2\,\varepsilon - 1}\,e^{\frac{\varepsilon t^2}{2k^2}+2\ell-1}\,\frac{t^2}{k^2}$$

by (4.16). In particular

$$\Lambda(t) \leq \frac{\varepsilon}{2\,\varepsilon - 1}\,e^{\frac{\varepsilon M^2}{2}+2\ell-1}\,\frac{t^2}{k^2} \tag{4.18}$$

for any $\varepsilon > 1/2$, $M > 0$, $\ell > 0$ and $0 \leq t \leq M\,k$.

4. From (4.17) and (4.18) it follows that

$$\Lambda(t) \leq \frac{1}{2\alpha}\,\sup(\ell, L)\left(\frac{4}{p}\,\frac{2}{M^2} + \frac{p}{\ell}\right)t^2 \tag{4.19}$$

for all $t \geq 0$ if M satisfies

$$\frac{\varepsilon}{2\,\varepsilon - 1}\,e^{\frac{\varepsilon M^2}{2}+2\ell-1} = \frac{2\,\ell}{p\,M^2} + \frac{p}{4}\cdot$$

We try to minimize the coefficient of t^2 in (4.19) by choosing ε, p and ℓ (hence M).

5. We actually let $\varepsilon > 1/2$ and $p \geq 1$ be fixed and we minimize over ℓ. For notational convenience we let $y = M^2/2$ and we try to minimize $\dfrac{4}{p}\,\dfrac{1}{y} + \dfrac{p}{\ell}\cdot$ Letting

$$f(\ell, y) = \frac{\varepsilon}{2\,\varepsilon - 1}\,e^{\varepsilon y + 2\ell - 1} - \frac{\ell}{p\,y} - \frac{p}{4}$$

for any $\ell > 0$, there exists a unique $y = y(\ell)$ such that $f(l, y(l)) = 0$ since $\dfrac{\partial f}{\partial y}(\ell, y) > 0$. Moreover $\dfrac{\partial y}{\partial \ell}(\ell) < 0$ for $\ell > \dfrac{1}{2}\cdot$

Then the function

$$g(\ell) = \frac{4}{p}\,\frac{1}{y(\ell)} + \frac{p}{\ell}\,\left(= \frac{4\,\varepsilon}{2\,\varepsilon - 1}\,\frac{1}{\ell}\,e^{\varepsilon y(\ell)+2\ell-1}\right)$$

has a unique minimum on $\ell > \dfrac{1}{2}$, achieved for ℓ such that $8\,\ell^2 - 4\,\ell - \varepsilon\,p^2\,y^2(\ell) = 0$. Since moreover $f(\ell, y(\ell)) = 0$, this minimal ℓ satisfies the equation $F(\ell) = 0$ where

$$F(\ell) = \frac{\varepsilon}{2\varepsilon - 1}\,e^{\frac{2}{p}\sqrt{\varepsilon}\,\sqrt{2\ell^2-\ell}+2\ell-1} - \frac{\ell\,\sqrt{\varepsilon}}{2\,\sqrt{2\ell^2 - \ell}} - \frac{p}{4}\cdot$$

This equation $F(\ell) = 0$ has a unique solution on $\ell > \dfrac{1}{2}$ since $F'(\ell) > 0$.

6. As a constant C one can choose

$$C = \Big(\frac{2\sqrt{\varepsilon}}{\sqrt{2\ell^2 - \ell}} + \frac{p}{\ell}\Big)^{1/2} \Big(\frac{2}{\alpha} \sup(\ell, L)\Big)^{1/2}$$

with $\ell > \dfrac{1}{2}$ solution to $F(\ell) = 0$.

We finally try to optimize over $p \geq 1$ and $\varepsilon > \dfrac{1}{2}\cdot$ For instance, for $p = 3$ and $\varepsilon = \dfrac{3}{2}$ we obtain $\ell = 0.677...$ which gives

$$C = 3.07 \Big(\frac{2}{\alpha} \sup(0.7, L)\Big)^{1/2}.$$

This concludes the argument. $\square$

Let us note that this proof is more direct than the one given above, but it is specific to this transportation inequality, and in the end leads to a slightly worse constant C (by the numerical factor 3.07).

Chapitre 5

Inégalités de concentration pour des variables dépendantes

Ce chapitre correspond à l'article [15] écrit en collaboration avec Gordon Blower et soumis pour publication.

Nous considérons dans ce chapitre des processus stochastiques à temps discret et à valeurs dans un espace polonais abstrait, et donnons des conditions suffisantes sur la distribution initiale et les mesures de transition pour que la loi jointe du processus vérifie des inégalités de concentration gaussienne ou de transport, ou encore une inégalité de Sobolev logarithmique dans le cas où l'espace d'état est l'espace euclidien $\mathbb{R}^m$. Dans de nombreux cas l'ordre de grandeur des constantes obtenues est optimal en le nombre de variables aléatoires considérées. Dans certains cas ces constantes sont même indépendantes du nombre de variables, étendant ainsi à des variables faiblement dépendantes des résutats connus pour des variables indépendantes.

Introduction

Given a complete and separable metric space (X, d), $\mathrm{Prob}(X)$ denotes the space of Radon probability measures on X, equipped with the (narrow) weak topology. We say that $\mu \in \mathrm{Prob}\,(X)$ satisfies a *Gaussian concentration inequality* $GC(\kappa)$ with constant κ on (X, d) if

$$\int_X \exp\bigl(tF(x)\bigr)\mu(dx) \leq \exp\Bigl(t\int_X F(x)\,\mu(dx) + \kappa\, t^2/2\Bigr) \qquad (t \in \mathbb{R})$$

holds for all 1-Lipschitz functions $F : (X, d) \to \mathbb{R}$ (see [17]). Recall that a function $g : (\Omega_1, d_1) \to (\Omega_2, d_2)$ between metric spaces is *L-Lipschitz* if $d_2(g(x), g(y)) \leq L\, d_1(x, y)$ holds for all $x, y \in \Omega_1$, and we call the infimum of such L the *Lipschitz seminorm* of g.

For $k \geq 1$ and $x_1, \ldots, x_k$ in X, we let $x^{(k)} = (x_1, \ldots, x_k) \in X^k$ and, given $1 \leq s < \infty$, we equip the product space X^k with the metric $d^{(s)}$ defined by $d^{(s)}(x^{(k)}, y^{(k)}) = (\sum_{j=1}^k d(x_j, y_j)^s)^{1/s}$ for $x^{(k)}$ and $y^{(k)}$ in X^k.

Now let $(\xi_j)_{j=1}^n$ be a stochastic process with state space X. The first aim of this paper is to obtain concentration inequalities for the joint distribution $P^{(n)}$ of $\xi^{(n)} = (\xi_1, \ldots, \xi_n)$, under hypotheses on the initial distribution $P^{(1)}$ of ξ_1 and the conditional distributions $p_k(. \mid x^{(k-1)})$ of ξ_k given $\xi^{(k-1)}$; we recall that $P^{(n)}$ is given by

$$P^{(n)}(dx^{(n)}) = p_n(dx_n \mid x^{(n-1)}) \ldots p_2(dx_2 \mid x_1)\, P^{(1)}(dx_1).$$

If the $(\xi_j)_{j=1}^n$ are mutually independent, and the distribution of each ξ_j satisfies $GC(\kappa)$, then $P^{(n)}$ on $(X^n, d^{(1)})$ is the product of the marginal distributions, and inherits $GC(n\kappa)$ from its marginal distributions by a simple 'tensorization' argument. A similar result also applies to product measures for the transportation and logarithmic Sobolev inequalities which we consider later ; see [68, 106]. To obtain concentration inequalities for $P^{(n)}$ when (ξ_j) are weakly dependent, we impose additional restrictions on the coupling between the variables, expressed in terms of Wasserstein distances which are defined as follows.

Given $1 \leq s < \infty$, $\mathrm{Prob}_s(X)$ denotes the subspace of $\mathrm{Prob}(X)$ consisting of ν such that $\int_X d(x_0, y)^s\, \nu(dy)$ is finite for some or equivalently all $x_0 \in X$. Then we define the *Wasserstein distance* of order s between μ and ν in $\mathrm{Prob}_s(X)$ by

$$W_s(\mu, \nu) = \inf_\pi \Big(\iint_{X \times X} d(x, y)^s\, \pi(dx\, dy) \Big)^{1/s} \tag{5.1}$$

where $\pi \in \mathrm{Prob}_s\,(X \times X)$ has marginals $\pi_1 = \mu$ and $\pi_2 = \nu$. Then W_s defines a metric on $\mathrm{Prob}_s(X)$, which in turn becomes a complete and separable metric space (see [93, 111]).

In section 5.2 we obtain the following result for time-homogeneous Markov chains.

Theorem 5.1. *Let $(\xi_j)_{j=1}^n$ be an homogeneous Markov process with state space X, initial distribution $P^{(1)}$ and transition measure $p(. \mid x)$. Suppose that there exist constants κ_1 and L such that :*

(i) $P^{(1)}$ and $p(. \mid x)$ $(x \in X)$ satisfy $GC(\kappa_1)$ on (X, d) ;

(ii) $x \mapsto p(. \mid x)$ is L-Lipschitz $(X, d) \to (\mathrm{Prob}_1\,(X), W_1)$.

Then the joint law $P^{(n)}$ of $(\xi_1, \ldots, \xi_n)$ satisfies $GC(\kappa_n)$ on $(X^n, d^{(1)})$, where

$$\kappa_n = \kappa_1 \sum_{m=1}^n \Big(\sum_{k=0}^{m-1} L^k \Big)^2.$$

In Example 5.15 we demonstrate sharpness of these constants by providing for each value of L a process such that κ_n has optimal growth in n.

Concentration inequalities are an instance of the wider class of transportation inequalities, which bound the transportation cost by the relative entropy. We recall the definitions.

Let ν and μ be in $\mathrm{Prob}(X)$, where ν is absolutely continuous with respect to μ, and let $d\nu/d\mu$ be the Radon–Nikodym derivative. Then we define the *relative entropy* of ν with respect to μ by

$$\mathrm{H}(\nu \mid \mu) = \int_X \log \frac{d\nu}{d\mu}\, d\nu\,;$$

note that $0 \leq \mathrm{H}(\nu \mid \mu) \leq \infty$ by Jensen's inequality. By convention we let $\mathrm{H}(\nu \mid \mu) = \infty$ if ν is not absolutely continuous with respect to μ.

Given $1 \leq s < \infty$, we say that $\mu \in \mathrm{Prob}_s(X)$ satisfies a *transportation inequality* $T_s(\alpha)$ for cost function $d(x,y)^s$, with constant α, if

$$W_s(\nu,\mu) \leq \Big(\frac{2}{\alpha}\mathrm{H}(\nu \mid \mu)\Big)^{1/2}$$

for all $\nu \in \mathrm{Prob}_s(X)$.

Marton [77] introduced T_2 as 'distance-divergence' inequalities in the context of information theory ; subsequently Talagrand [106] showed that the standard Gaussian distribution on $\mathbb{R}^m$ satisfies $T_2(1)$. Bobkov and Götze showed in [17] that $GC(\kappa)$ is equivalent to $T_1(1/\kappa)$; their proof used the Kantorovich–Rubinstein duality result, that

$$W_1(\mu,\nu) = \sup_f \Big\{\int_X f(x)\mu(dx) - \int_X f(y)\,\nu(dy)\Big\}$$

where $\mu,\nu \in \mathrm{Prob}_1(X)$ and f runs over the set of 1-Lipschitz functions $f : X \to \mathbb{R}$. A $\nu \in \mathrm{Prob}(X)$ satisfies a T_1 inequality if and only if ν admits a square-exponential moment ; that is, $\int_X \exp(\beta d(x,y)^2)\,\nu(dx)$ is finite for some $\beta > 0$ and some, and thus all, $y \in X$; see [23, 46] for detailed statements. Moreover, since $T_s(\alpha)$ implies $T_r(\alpha)$ for $1 \leq r \leq s$ by Hölder's inequality, transportation inequalities are a tool for proving and strengthening concentration inequalities ; they are also related to the Gaussian isoperimetric inequality as in [14]. For applications to empirical distributions in statistics, see [81].

Returning to weakly dependent $(\xi_j)_{j=1}^n$ with state space X, we obtain transportation inequalities for the joint distribution $P^{(n)}$, under hypotheses on $P^{(1)}$ and the conditional distributions. Djellout, Guillin and Wu [46] developed Marton's coupling method [78, 79] to prove $T_s(\alpha)$ for $P^{(n)}$ under various mixing or contractivity conditions ; see also [96], or [23] where the conditions are expressed solely in terms of exponential moments. We extend these results in sections 5.1 and 5.2 below, thus obtaining a strengthened dual form of Theorem 5.1.

Theorem 5.2. *Let $(\xi_j)_{j=1}^n$ be an homogeneous Markov process with state space X, initial distribution $P^{(1)}$ and transition measure $p(.\mid x)$. Suppose that there exist constants $1 \leq s \leq 2$, $\alpha > 0$ and $L \geq 0$ such that :*

(i) $P^{(1)}$ and $p(.\mid x)$ $(x \in X)$ satisfy $T_s(\alpha)$;

(ii) $x \mapsto p(.\mid x)$ is L-Lipschitz $(X,d) \to (Prob_s(X), W_s)$.

Then the joint distribution $P^{(n)}$ of $(\xi_1, \dots, \xi_n)$ satisfies $T_s(\alpha_n)$, where

$$\alpha_n = \begin{cases} n^{1-(2/s)}(1 - L^{1/s})^2 \alpha & \text{if} \quad L < 1, \\ e^{(2/s)-2}(n^{-(2/s)-1}\alpha & \text{if} \quad L = 1, \\ \Big(\dfrac{L-1}{e^{s-1}L^n}\Big)^{2/s} \dfrac{\alpha}{n+1} & \text{if} \quad L > 1; \end{cases}$$

in particular α_n is independent of n for $s = 2$ when $L < 1$.

Our general transportation Theorem 5.4 will involve processes that are not necessarily Markovian, but satisfy some a hypothesis related to Dobrushin–Shlosman's mixing condition (see [79, Definition 2] for instance). When $X = \mathbb{R}^m$, we shall also present some more computable version of hypothesis (ii) in Proposition 5.5, and later consider a stronger functional inequality.

A probability measure μ on $\mathbb{R}^m$ satisfies the *logarithmic Sobolev inequality* $LSI(\alpha)$ with constant $\alpha > 0$ if

$$\int_{\mathbb{R}^m} f^2 \log\Big(f^2 / \int_{\mathbb{R}^m} f^2 \, d\mu\Big) d\mu \le \frac{2}{\alpha} \int_{\mathbb{R}^m} \|\nabla f\|_{\ell^2}^2 \, d\mu$$

holds for all $f \in L^2(d\mu)$ that have distributional gradient $\nabla f \in L^2(d\mu; \mathbb{R}^m)$. Given $(a_k) \in \mathbb{R}^m$, let $\|(a_k)\|_{\ell^s} = (\sum_{k=1}^m |a_k|^s)^{1/s}$ for $1 \le s < \infty$, and $\|(a_k)\|_{\ell^\infty} = \sup_{1 \le k \le m} |a_k|$.

The connection between the various inequalities is summarized by

$$LSI(\alpha) \Rightarrow T_2(\alpha) \Rightarrow T_1(\alpha) \Leftrightarrow GC(1/\alpha); \tag{5.2}$$

see [17, 88, 111]. Conversely, Otto and Villani showed that if $\mu(dx) = e^{-V(x)}\, dx$ satisfies $T_2(\alpha)$ where $V : \mathbb{R}^m \to \mathbb{R}$ is convex, then μ also satisfies $LSI(\alpha/4)$ (see [16, 88, 111]; but this converse is not generally true, as a counter-example in [39] shows.

Gross [58] proved that the standard Gaussian probability measure on $\mathbb{R}^m$ satisfies $LSI(1)$. More generally, Bakry and Emery [8] showed that if V is twice continuously differentiable, with $\operatorname{Hess} V \ge \alpha I_m$ on $\mathbb{R}^m$ for some $\alpha > 0$, then $\mu(dx) = e^{-V(x)}\, dx$ satisfies $LSI(\alpha)$; see for instance [112] for extensions to this result. Whereas Bobkov and Götze [17] characterized in terms of their cumulative distribution functions those $\mu \in \mathrm{Prob}(\mathbb{R})$ that satisfy $LSI(\alpha)$ for some α, there is no known geometrical characterization of such probability measures on $\mathbb{R}^m$ when $m > 1$.

Our main Theorem 5.10 gives a sufficient condition for the joint law of a weakly dependent process with state space $\mathbb{R}^m$ to satisfy LSI. In section 5.5 we deduce the following for distributions of time-homogeneous Markov processes. Let $\partial/\partial x$ denote the gradient with respect to $x \in \mathbb{R}^m$.

Theorem 5.3. *Let $(\xi_j)_{j=1}^n$ be an homogeneous Markov process with state space $\mathbb{R}^m$, initial distribution $P^{(1)}$ and transition measure $p(dy|x) = e^{-u(x,y)}dy$. Suppose that there exist constants $\alpha > 0$ and $L \ge 0$ such that :*

(i) $P^{(1)}$ *and* $p(.\mid x)$ $(x \in \mathbb{R}^m)$ *satisfy* $LSI(\alpha)$ *;*

(ii) u *is twice continuously differentiable and the off-diagonal blocks of its Hessian matrix satisfy*

$$\Big\| \frac{\partial^2 u}{\partial x \partial y} \Big\| \leq L$$

as operators $(\mathbb{R}^m, \ell^2) \to (\mathbb{R}^m, \ell^2)$.
Then the joint law $P^{(n)}$ *of the first* n *variables* $(\xi_1, \ldots, \xi_n)$ *satisfies* $LSI(\alpha_n)$*, where*

$$\alpha_n = \begin{cases} \dfrac{(\alpha - L)^2}{\alpha} & \text{if} \quad L < \alpha, \\ \dfrac{\alpha}{n(n+1)(e-1)} & \text{if} \quad L = \alpha, \\ \Big(\dfrac{\alpha}{L}\Big)^{2n} \dfrac{L^2 - \alpha^2}{\alpha e(n+1)} & \text{if} \quad L > \alpha; \end{cases}$$

in particular α_n *is independent of* n *when* $L < \alpha$.

The plan of the paper is as follows. In section 5.1 we state and prove our results on transportation inequalities, which imply Theorem 5.2, and in section 5.2 we deduce Theorem 5.1. In section 5.3 we prove $LSI(\alpha)$ for the joint distribution of ARMA processes, with α independent of the size of the sample. In section 5.4 we obtain a more general LSI, which we express in a simplified form for Markov processes in section 5.5. Explicit examples in section 5.5 show that several of our results have optimal growth of the constants with respect to n as $n \to \infty$, and that the hypotheses are computable and realistic.

5.1 Transportation inequalities

Let $(\xi_k)_{k=1}^n$ be a stochastic process with state space X, let $p_k(.\mid x^{(k-1)})$ denote the transition measure between the states at times $k-1$ and k, and let $P^{(n)}$ be the joint distribution of $\xi^{(n)}$. Our main result of this section is a transportation inequality.

Theorem 5.4. *Let* $1 \leq s \leq 2$*, and suppose that there exist* $\alpha_1 > 0$ *and* $M \geq \rho_\ell \geq 0$ $(\ell = 1, \ldots, n)$ *such that :*

(i) $P^{(1)}$ *and* $p_k(.\mid x^{(k-1)})$ $(k = 2, \ldots, n;\ x^{(k-1)} \in X^{k-1})$ *satisfy* $T_s(\alpha)$ *on* (X, d) *;*

(ii) $x^{(k-1)} \mapsto p_k(.\mid x^{(k-1)})$ *is Lipschitz as a map* $(X^{k-1}, d^{(s)}) \to (Prob_s(X), W_s)$ *for* $k = 2, \ldots, n$*, in the sense that*

$$W_s\big(p_k(.\mid x^{(k-1)}), p_k(.\mid y^{(k-1)})\big)^s \leq \sum_{j=1}^{k-1} \rho_{k-j}\, d(x_j, y_j)^s \qquad (x^{(k-1)}, y^{(k-1)} \in X^{k-1}).$$

Then $P^{(n)}$ *satisfies the transportation inequality* $T_s(\alpha_n)$ *where*

$$\alpha_n = \alpha \left(\frac{(n\, e)^{1-s} M}{(1+M)^n} \right)^{2/s}.$$

Suppose further that

(iii) $\sum_{j=1}^{n} \rho_j \leq R$.

Then the joint distribution $P^{(n)}$ *satisfies* $T_s(\alpha_n)$ *where*

$$\alpha_n = \begin{cases} n^{1-(2/s)}(1 - R^{1/s})^2\alpha & \text{if} \quad R < 1, \\ e^{(2/s)-2}(n+1)^{-(2/s)-1} & \text{if} \quad R = 1, \\ \Big(\dfrac{R-1}{e^{s-1}R^n}\Big)^{2/s} \dfrac{\alpha}{n+1} & \text{if} \quad R > 1. \end{cases}$$

In hypothesis (iii), the sequence $(\rho_k)_{k=1}^{n-1}$ measures the extent to which the distribution of ξ_n depends upon the previous $\xi_{n-1}, \xi_{n-2}, \dots$; so in most examples $(\rho_k)_{k=1}^{n-1}$ is decreasing.

A version of Theorem 5.4 was obtained by Djellout, Guillin and Wu, but with an explicit constant only when $R < 1$; see [46, Theorem 2.5 and Remark 2.9]. Theorem 5.4 also improves upon section 4 of [23], where the assumptions were written in terms of moments of the considered measures.

The Monge–Kantorovich transportation problem involves finding, for given $\mu, \nu \in \mathrm{Prob}(X)$, an optimal transportation strategy in (5.1), namely a π that minimises the transportation cost ; a compactness and semi-continuity argument ensures that, for suitable cost functions, there always exists such a π. We recall that, given $\mu \in \mathrm{Prob}(X)$, another Polish space Y and a continuous function $\varphi : X \to Y$, the measure *induced* from μ by φ is the unique $\nu \in \mathrm{Prob}(Y)$ such that

$$\int_Y f(y)\nu(dy) = \int_X f(\varphi(x))\mu(x)$$

for all bounded and continuous $f : X \to \mathbb{R}$. Brenier [28] and McCann [82] showed that if μ and ν belong to $\mathrm{Prob}_2(\mathbb{R}^m)$, and if moreover μ is absolutely continuous with respect to Lebesgue measure, then there exists a convex function $\Phi : \mathbb{R}^m \to \mathbb{R}$ such that the gradient $\varphi = \nabla\Phi$ induces μ from ν and gives the unique solution to the Monge–Kantorovich transportation problem for $s = 2$, in the sense that

$$\int_{\mathbb{R}^m} \|\nabla\Phi(x) - x\|_{\ell^2}^2 \, \mu(dx) = W_2(\mu, \nu)^2.$$

Further extensions of this result were obtained by Gangbo and McCann [52] for $1 < s \leq 2$, by Ambrosio and Pratelli [4] for $s = 1$, and by McCann [83] in the context of compact and connected C^3-smooth Riemannian manifolds that are without boundary (see also[111]).

Proof of Theorem 5.4. In order to give an explicit solution in a case of importance, we first suppose that $X = \mathbb{R}^m$ and that $P^{(1)}$ and $p_j(dx_j \mid x^{(j-1)})$ $(j = 2, \dots, n)$ are all absolutely continuous with respect to Lebesgue measure. Then let $Q^{(n)} \in \mathrm{Prob}_s(\mathbb{R}^{nm})$ be of finite relative entropy with respect to $P^{(n)}$. Let $Q^{(j)}(dx^{(j)})$ be the marginal distribution of $x^{(j)} \in \mathbb{R}^{jm}$ with respect to $Q^{(n)}(dx^{(n)})$, and disintegrate $Q^{(n)}$ in terms of conditional probabilities, according to

$$Q^{(j)}(dx^{(j)}) = q_j(dx_j \mid x^{(j-1)})Q^{(j-1)}(dx^{(j-1)}).$$

In particular $q_j(. \mid x^{(j-1)})$ is absolutely continuous with respect to $p_j(. \mid x^{(j-1)})$ and hence with respect to Lebesgue measure, for $Q^{(j-1)}$ almost every $x^{(j-1)}$. A standard computation ensures that

$$\begin{aligned} \mathrm{H}\left(Q^{(n)} \mid P^{(n)}\right) &= \mathrm{H}\left(Q^{(1)} \mid P^{(1)}\right) \\ &+ \sum_{j=2}^{n} \int_{\mathbb{R}^{(j-1)m}} \mathrm{H}\left(q_j(\,.\, \mid x^{(j-1)}) \mid p_j(\,.\, \mid x^{(j-1)})\right) Q^{(j-1)}(dx^{(j-1)}). \end{aligned} \tag{5.3}$$

When the hypothesis (i) of Theorem 5.4 holds for some $1 < s \leq 2$, it also holds for $s = 1$. Consequently, by the Bobkov–Götze theorem, $P^{(1)}$ and $p_j(dx_j \mid x^{(j-1)})$ satisfy $GC(\kappa)$ for $\kappa = 1/\alpha$, and then one can check that there exists $\varepsilon > 0$ such that

$$\int_{\mathbb{R}^m} \exp(\varepsilon \|x^{(1)}\|_{\ell^2}^2) P^{(1)}(dx^{(1)}) < \infty$$

and likewise for p_j; compare with Herbst's theorem [111, p. 280], and [17, 46]. Hence $Q^{(1)}$ and $q_j(dx_j \mid x^{(j-1)})$ for $Q^{(j-1)}$ almost every $x^{(j-1)}$ have finite second moments, since by Young's inequality

$$\int_{\mathbb{R}^m} \varepsilon \|x^{(n)}\|_{\ell^2}^2 \, Q^{(1)}(dx^{(1)}) \leq \mathrm{H}(Q^{(1)} \mid P^{(1)}) + \log \int_{\mathbb{R}^m} \exp\bigl(\varepsilon \|x^{(1)}\|_{\ell^2}^2\bigr) P^{(1)}(dx^{(1)}) < \infty$$

and likewise with q_j and p_j in place of $Q^{(1)}$ and $P^{(1)}$ respectively.

Let $\theta_1 : \mathbb{R}^m \to \mathbb{R}^m$ be an optimal transportation map that induces $P^{(1)}(dx_1)$ from $Q^{(1)}(dx_1)$; then for $Q^{(1)}$ every each x_1, let $x_2 \mapsto \theta_2(x_1, x_2)$ induce $p_2(dx_2 \mid \theta_1(x_1))$ from $q_2(dx_2 \mid x_1)$ optimally; hence $\Theta^{(2)} : \mathbb{R}^{2m} \to \mathbb{R}^{2m}$, defined by $\Theta^{(2)}(x_1, x_2) = (\theta_1(x_1), \theta_2(x_1, x_2))$ on a certain set of full $Q^{(2)}$ measure, induces $P^{(2)}$ from $Q^{(2)}$. Generally, having constructed $\Theta^{(j)} : \mathbb{R}^{jm} \to \mathbb{R}^{jm}$, we let $x_{j+1} \mapsto \theta_{j+1}(x^{(j)}, x_{j+1})$ be an optimal transportation map that induces $p_{j+1}(dx_{j+1} \mid \Theta^{(j)}(x^{(j)}))$ from $q_{j+1}(dx_{j+1} \mid x^{(j)})$, for all $x^{(j)}$ in a certain set of full $Q^{(j)}$ measure; then we let $\Theta^{(j+1)} : \mathbb{R}^{(j+1)m} \to \mathbb{R}^{jm} \times \mathbb{R}^m$ be the map defined by

$$\Theta^{(j+1)}(x^{(j+1)}) = (\Theta^{(j)}(x^{(j)}), \theta_{j+1}(x^{(j+1)}))$$

on a set of full $Q^{(j+1)}$ measure. In particular $\Theta^{(j+1)}$ induces $P^{(j+1)}$ from $Q^{(j+1)}$, in the style of Kneser.

This transportation strategy may not be optimal, nevertheless it gives the bound

$$W_s(Q^{(n)}, P^{(n)})^s \leq \int_{\mathbb{R}^{nm}} \|\Theta^{(n)}(x^{(n)}) - x^{(n)}\|_{\ell^s}^s Q^{(n)}(dx^{(n)}) = \sum_{k=1}^{n} d_k \tag{5.4}$$

by the recursive definition of $\Theta^{(n)}$, where we have let

$$d_k = \int_{\mathbb{R}^{km}} \|\theta_k(x^{(k)}) - x_k\|_{\ell^s}^s Q^{(k)}(dx^{(k)}) \qquad (k = 1, \ldots, n).$$

However, the transportation at step k is optimal by construction, so

$$d_k = \int_{\mathbb{R}^{(k-1)m}} W_s\big(p_k(\,.\mid \Theta^{(k-1)}(x^{(k-1)})), q_k(\,.\mid x^{(k-1)})\big)^s Q^{(k-1)}(dx^{(k-1)}). \tag{5.5}$$

Given $a, b > 0$, $1 \le s \le 2$ and $\gamma > 1$, we have $(a+b)^s \le (\gamma/(\gamma-1))^{s-1}a^s + \gamma^{s-1}b^s$. Hence by the triangle inequality, the expression (5.5) is bounded by

$$\Big(\frac{\gamma}{\gamma-1}\Big)^{s-1}\int_{\mathbb{R}^{(k-1)m}} W_s\big(p_k(\,.\mid x^{(k-1)}), q_k(\,.\mid x^{(k-1)})\big)^s Q^{(k-1)}(dx^{(k-1)})$$
$$+\gamma^{s-1}\int_{\mathbb{R}^{(k-1)m}} W_s(p_k(\,.\mid \Theta^{(k-1)}(x^{(k-1)})), p_k(\,.\mid x^{(k-1)}))^s Q^{(k-1)}(dx^{(k-1)}). \tag{5.6}$$

By hypothesis (i) and then Hölder's inequality, we bound the first integral in (5.6) by

$$h_k = \Big(\frac{\gamma}{\gamma-1}\Big)^{s-1}\Big(\frac{2}{\alpha}\Big)^{s/2}\Big(\int_{\mathbb{R}^{(k-1)m}} \mathrm{H}(q_k \mid p_k)\, dQ^{(k-1)}\Big)^{s/2}.$$

Meanwhile, on account of hypothesis (ii) the second integral in (5.6) is bounded by

$$\gamma^{s-1}\int_{\mathbb{R}^{(k-1)d}} \sum_{j=1}^{k-1} \rho_{k-j}\big\|\theta_j(x^{(j)}) - x_j\big\|^s Q^{(k-1)}(dx^{(k-1)}) = \gamma^{s-1}\sum_{j=1}^{k-1}\rho_{k-j}d_j,$$

and when we combine these contributions to (5.6) we have

$$d_k \le h_k + \gamma^{s-1}\sum_{j=1}^{k-1}\rho_{k-j}\, d_j. \tag{5.7}$$

In the case when the ρ_ℓ are merely bounded by M, one can prove by induction that

$$d_k \le h_k + \gamma^{s-1}M\sum_{j=1}^{k-1} h_j(1+\gamma^{s-1}M)^{k-1-j},$$

so that

$$\sum_{k=1}^n d_k \le \sum_{j=1}^n h_j(1+\gamma^{s-1}M)^{n-j} \le \Big(\sum_{j=1}^n h_j^{2/s}\Big)^{s/2}\Big(\sum_{j=1}^n (1+\gamma^{s-1}M)^{2(n-j)/(2-s)}\Big)^{(2-s)/2}$$

by Hölder's inequality. The first sum on the right-hand side is

$$\Big(\sum_{j=1}^n h_j^{2/s}\Big)^{s/2} = \Big(\frac{\gamma}{\gamma-1}\Big)^{s-1}\Big(\frac{2}{\alpha}\Big)^{s/2}\mathrm{H}\,(Q^{(n)} \mid P^{(n)})^{s/2}$$

by (5.3). Finally, setting $\gamma = 1 + 1/n$, we obtain by (5.4) the stated result

$$W_s(Q^{(n)}, P^{(n)})^s \le \Big(\frac{2}{\alpha}\Big)^{s/2}\frac{(1+M)^n}{M}\,(n\,e)^{s-1}\,\mathrm{H}(Q^{(n)} \mid P^{(n)})^{s/2}.$$

(iii) Invoking the further hypothesis (iii), we see that $T_m = \sum_{j=1}^m d_j$ satisfies on account of (5.7) the recurrence relation

$$T_{m+1} \le \sum_{j=1}^{m+1} h_j + \gamma^{s-1} R\, T_m,$$

which enables us to use Hölder's inequality again and bound T_n by

$$\begin{aligned}\sum_{k=1}^n \Big(\sum_{j=1}^k h_j\Big)(\gamma^{s-1}R)^{n-k} &= \sum_{j=1}^n h_j \sum_{\ell=0}^{n-j} (\gamma^{s-1}R)^\ell \\ &\le \Big(\sum_{j=1}^n h_j^{2/s}\Big)^{s/2} \Big(\sum_{j=1}^n \Big(\sum_{\ell=0}^{n-j} (\gamma^{s-1}R)^\ell\Big)^{2/(2-s)}\Big)^{(2-s)/2}\end{aligned}$$

for $1 \le s < 2$. By (5.4) and the definition of T_n this leads to

$$\begin{aligned} W_s(Q^{(n)}, P^{(n)})^s &\le \Big(\frac{\gamma}{\gamma-1}\Big)^{s-1} \Big(\sum_{m=1}^n \Big(\sum_{\ell=0}^{m-1} (\gamma^{s-1}R)^\ell\Big)^{2/(2-s)}\Big)^{(2-s)/2} \Big(\frac{2}{\alpha}\,\mathrm{H}(Q^{(n)} \mid P^{(n)})\Big)^{s/2} && (5.8)\\ &\le \Big(\frac{\gamma}{\gamma-1}\Big)^{s-1} n^{1-s/2} \sum_{\ell=0}^{n-1} (\gamma^{s-1}R)^\ell \Big(\frac{2}{\alpha}\,\mathrm{H}(Q^{(n)} \mid P^{(n)})\Big)^{s/2}; && (5.9)\end{aligned}$$

this also holds for $s = 2$. Finally we select γ according to the value of R to make the bound (5.9) precise. When $R < 1$, we let $\gamma = R^{-1/s} > 1$, so that $\gamma^{s-1}R = R^{1/s} < 1$, and we deduce the transportation inequality

$$W_s(Q^{(n)}, P^{(n)})^s \le \Big(\frac{2}{\alpha}\Big)^{s/2} \frac{n^{1-s/2}}{(1-R^{1/s})^s} \mathrm{H}(Q^{(n)} \mid P^{(n)})^{s/2}.$$

When $R \ge 1$, we let $\gamma = 1 + 1/n$ to obtain the transportation inequality

$$W_s(Q^{(n)}, P^{(n)})^s \le \Big(\frac{2}{\alpha}\Big)^{s/2} (n+1)^{s-1} n^{1-s/2} \left(\frac{(1+1/n)^{n(s-1)}R^n - 1}{(1+1/n)^{s-1}R - 1}\right) \mathrm{H}(Q^{(n)} \mid P^{(n)})^{s/2},$$

which leads to the stated result by simple analysis, and completes the proof when $X = \mathbb{R}^m$.

For typical Polish spaces (X, d), we cannot rely on the existence of optimal maps, but we can use a less explicit inductive approach to construct the transportation strategy, as in [46]. Given $j = 1, \dots, n-1$, assume that $\pi^{(j)} \in \mathrm{Prob}(X^{2j})$ has marginals $Q^{(j)}(dx^{(j)})$ and $P^{(j)}(dy^{(j)})$ and satisfies

$$W_s(Q^{(j)}, P^{(j)})^s \le \int_{X^{2j}} \sum_{k=1}^j d(x_k, y_k)^s \pi^{(j)}(dx^{(j)} dy^{(j)}).$$

Then, for each $(x^{(j)}, y^{(j)}) \in X^{2j}$, let $\sigma_{j+1}(\,.\,\mid x^{(j)}, y^{(j)}) \in \mathrm{Prob}(X^2)$ be an optimal transportation strategy that has marginals $q_{j+1}(dx_{j+1} \mid x^{(j)})$ and $p_{j+1}(dy_{j+1} \mid y^{(j)})$ and that satisfies

$$W_s(q_{j+1}(\,.\,\mid x^{(j)}), p_{j+1}(\,.\,\mid y^{(j)}))^s = \int_{X^2} d(x_{j+1}, y_{j+1})^s\, \sigma_{j+1}(dx_{j+1}\, dy_{j+1} \mid x^{(j)}, y^{(j)}).$$

Now we let

$$\pi^{(j+1)}(dx^{(j+1)}dy^{(j+1)}) = \sigma_{j+1}(dx_{j+1}\, dy_{j+1} \mid x^{(j)}, y^{(j)})\, \pi^{(j)}(dx^{(j)}dy^{(j)}),$$

which defines a probability on $X^{2(j+1)}$ with marginals $Q^{(j+1)}(dx^{(j+1)})$ and $P^{(j+1)}(dy^{(j+1)})$. This may not give an optimal transportation strategy; nevertheless, the recursive definition shows that

$$W_s(Q^{(n)}, P^{(n)})^s \le \sum_{j=1}^{n} \int_{X^{2(j-1)}} W_s\big(q_j(\,.\,\mid x^{(j-1)}), p_j(\,.\,\mid y^{(j-1)})\big)^s \pi^{(j-1)}(dx^{(j-1)}dy^{(j-1)})$$

and one can follow the preceding proof from (5.4) onwards. □

Proof of Theorem 5.2. Under the hypotheses of Theorem 5.2, we can take $\rho_1 = L$ and $\rho_j = 0$ for $j = 2, \dots, n$, which satisfy Theorem 5.4 with $R = L$ in assumption (iii). □

The definition of W_s not being well suited to direct calculation, we now give a computable sufficient condition for hypothesis (ii) of Theorem 5.4 to hold with some constant coefficients ρ_ℓ when $(X, d) = (\mathbb{R}^m, \ell^s)$.

Proposition 5.5. *Let $u_j : \mathbb{R}^{jd} \to \mathbb{R}$ be a twice continuously differentiable function that has bounded second-order partial derivatives. Let $1 \le s \le 2$ and suppose further that :*

(i) $p_j(dx_j \mid x^{(j-1)}) = \exp(-u_j(x^{(j)}))\, dx_j$ satisfies $T_s(\alpha)$ for some $\alpha > 0$ and all $x^{(j-1)} \in \mathbb{R}^{(j-1)m}$;

(ii) there exists some real number M_s such that

$$\sup_{x^{(j-1)}} \int_{\mathbb{R}^m} \Big\| \Big(\frac{\partial u_j}{\partial x_k}\Big)_{k=1}^{j-1} \Big\|_{\ell^{s'}}^2 p_j(dx_j \mid x^{(j-1)}) = M_s,$$

where $1/s' + 1/s = 1$ and $\partial/\partial x_k$ denotes the gradient with respect to x_k.
Then $x^{(j-1)} \mapsto p_j(.\mid x^{(j-1)})$ is $\sqrt{(M_s/\alpha)}$-Lipschitz $(\mathbb{R}^{(j-1)m}, \ell^s) \to (Prob_s(\mathbb{R}^m), W_s)$.

Proof. Given $x^{(j-1)}, \bar{x}^{(j-1)} \in \mathbb{R}^{(j-1)m}$, we let $x^{(j-1)}(t) = (1-t)\bar{x}^{(j-1)} + t x^{(j-1)}$ $(0 \le t \le 1)$ be the straight-line segment that joins them, and we consider

$$f(t) = W_s\big(p_j(.\mid x^{(j-1)}(t)), p_j(.\mid \bar{x}^{(j-1)})\big);$$

then it suffices to show that $f : [0,1] \to \mathbb{R}$ is Lipschitz and to bound its Lipschitz seminorm.

By the triangle inequality and (i), we have

$$\Big(\frac{f(t+\delta) - f(t)}{\delta}\Big)^2$$

$$\leq \frac{1}{\delta^2} W_s\big(p_j(.\mid x^{(j-1)}(t+\delta)), p_j(.\mid x^{(j-1)}(t))\big)^2$$

$$\leq \frac{1}{\alpha\delta^2}\Big\{\mathrm{H}\big(p_j(.\mid x^{(j-1)}(t+\delta)) \mid p_j(.\mid x^{(j-1)}(t))\big) + \mathrm{H}\big(p_j(.\mid x^{(j-1)}(t)) \mid p_j(.\mid x^{(j-1)}(t+\delta))\big)\Big\}$$

$$= \frac{1}{\alpha\delta^2}\int_{\mathbb{R}^m} \big(u_j(x^{(j-1)}(t+\delta), x_j) - u_j(x^{(j-1)}(t), x_j)\big)$$
$$\big\{\exp(-u_j(x^{(j-1)}(t), x_j)) - \exp(-u_j(x^{(j-1)}(t+\delta), x_j))\big\}\, dx_j. \quad (5.10)$$

However, by the assumptions on u_j and the mean-value theorem, we have

$$u_j(x^{(j-1)}(t+\delta), x_j) - u_j(x^{(j-1)}(t), x_j)$$
$$= \delta \sum_{k=1}^{j-1} \Big\langle \frac{\partial u_j}{\partial x_k}(x^{(j-1)}(t), x_j), x_k - \bar{x}_k \Big\rangle + \frac{\delta^2}{2}\Big\langle \mathrm{Hess}\, u_j\, (x^{(j-1)} - \bar{x}^{(j-1)}), (x^{(j-1)} - \bar{x}^{(j-1)})\Big\rangle,$$

where $\mathrm{Hess}\, u_j$ is computed at some point between $(x^{(j-1)}, x_j)$ and $(\bar{x}^{(j-1)}, x_j)$ and is uniformly bounded. Proceeding in the same way for the other term (5.10), we obtain

$$\limsup_{\delta\to 0+}\Big(\frac{f(t+\delta) - f(t)}{\delta}\Big)^2 \leq \frac{1}{\alpha}\int_{\mathbb{R}^m}\Big(\sum_{k=1}^{j-1}\Big\langle \frac{\partial u_j}{\partial x_k}(x^{(j-1)}(t), x_j), x_k - \bar{x}_k\Big\rangle\Big)^2 p_j(dx_j \mid x^{(j-1)}(t)).$$

Hence by Hölder's inequality we have

$$\limsup_{\delta\to 0+} \frac{|f(t+\delta) - f(t)|}{\delta}$$
$$\leq \frac{1}{\sqrt{\alpha}}\bigg(\int_{\mathbb{R}^m}\Big(\sum_{k=1}^{j-1}\Big|\frac{\partial u_j}{\partial x_k}(x^{(j-1)}(t), x_j)\Big|^{s'}\Big)^{2/s'} p_j(dx_j \mid x^{(j-1)}(t))\bigg)^{1/2} \|x^{(j-1)} - \bar{x}^{(j-1)}\|_{\ell^s}$$

for $1 < s \leq 2$, and likewise with obvious changes for $s = 1$. By assumption (ii) and Vitali's theorem, f is Lipschitz with constant $\sqrt{(M_s/\alpha)}\|x^{(j-1)} - \bar{x}^{(j-1)}\|_{\ell^s}$, as required. □

5.2 Concentration inequalities for weakly dependent sequences

In terms of concentration inequalities, the dual version of Theorem 5.4 reads as follows.

Theorem 5.6. *Suppose that there exist $\kappa_1 > 0$ and $M \geq \rho_j \geq 0$ ($j = 1, \ldots, n$) such that :*
(i) $P^{(1)}$ and $p_k(.\mid x^{(k-1)})$ ($k = 2, \ldots, n$; $x^{(k-1)} \in X^{k-1}$) satisfy $GC(\kappa_1)$ on (X, d) ;
(ii) $x^{(k-1)} \mapsto p_k(.\mid x^{(k-1)})$ is Lipschitz as a map $(X^{k-1}, d^{(1)}) \to (Prob_1(X), W_1)$ for $k = 2, \ldots, n$, in the sense that

$$W_1\big(p_k(.\mid x^{(k-1)}), p_k(.\mid y^{(k-1)})\big) \leq \sum_{j=1}^{k-1} \rho_{k-j}\, d(x_j, y_j) \qquad (x^{(k-1)}, y^{(k-1)} \in X^{k-1}).$$

Then the joint law $P^{(n)}$ satisfies $GC(\kappa_n)$ on $(X^n, d^{(1)})$, where

$$\kappa_n = \kappa_1 \frac{(1+M)^{2n}}{M^2}.$$

Suppose moreover that
(iii) $\sum_{j=1}^n \rho_j \le R$.
Then $P^{(n)}$ satisfies $GC(\kappa_n(R))$ on $(X^n, d^{(1)})$, where

$$\kappa_n(R) = \kappa_1 \sum_{m=1}^{n} \Big(\sum_{k=0}^{m-1} R^k \Big)^2.$$

Proof of Theorem 5.6. This follows from the Bobkov–Götze theorem [17] and the bound (5.8) with $s = 1$ in the proof of Theorem 5.4. □

Alternatively, one can prove Theorem 5.6 directly by induction on the dimension, using the definition of GC.

Proof of Theorem 5.6. Under the hypotheses of Theorem 5.1, we can apply Theorem 5.6 with $\rho_1 = L$ and $\rho_j = 0$ for $j = 2, \dots, n$, which satisfy (iii). □

5.3 Logarithmic Sobolev inequalities for ARMA models

In this section we give logarithmic Sobolev inequalities for the joint law of the first n variables from two auto-regressive moving average processes. In both results we obtain constants that are independent of n, though the variables are not mutually independent, and we rely on the following general result which induces logarithmic Sobolev inequalities from one probability measure to another. For $d \ge 1$, let $\nu \in \mathrm{Prob}(\mathbb{R}^m)$ satisfy $LSI(\alpha)$, and let φ be a L-Lipschitz map from $(\mathbb{R}^m, \ell^2)$ into itself ; then, by the chain rule, the probability measure that is induced from ν by φ satisfies $LSI(\alpha/L^2)$. Our first application is the following.

Proposition 5.7. *Let Z_0 and Y_j $(j = 1, 2, \dots)$ be mutually independent random variables in $\mathbb{R}^m$, and let $\alpha > 0$ be a constant such that the distribution $P^{(0)}$ of Z_0 and the distribution of Y_j $(j = 1, 2, \dots)$ satisfy $LSI(\alpha)$.*
Then for any L-Lipschitz map $\Theta : \mathbb{R}^m \to \mathbb{R}^m$, the relation

$$Z_{j+1} = \Theta(Z_j) + Y_{j+1} \qquad (j = 0, 1, \dots) \tag{5.11}$$

determines a stochastic process such that, for any $n \ge 1$, the joint distribution $P^{(n-1)}$ of $(Z_j)_{j=0}^{n-1}$ satisfies $LSI(\alpha_n)$ where

$$\alpha_n = \begin{cases} (1-L)^2 \alpha & \text{if} \quad 0 \le L < 1, \\ \dfrac{\alpha}{n(n+1)(e-1)} & \text{if} \quad L = 1, \\ \dfrac{L-1}{L^n} \dfrac{\alpha}{e(n+1)} & \text{if} \quad L > 1. \end{cases}$$

Proof. For $(z_0, y_1, \dots, y_{n-1}) \in \mathbb{R}^{nm}$, let $\varphi_n(z_0, y_1, \dots, y_{n-1})$ be the vector $(z_0, \dots, z_{n-1})$, defined by the recurrence relation

$$z_{k+1} = \Theta(z_k) + y_{k+1} \qquad (k = 0, \dots, n-2). \tag{5.12}$$

Using primes to indicate another solution of (5.12), we deduce the following inequality from the Lipschitz condition on Θ :

$$\|z_{k+1} - z'_{k+1}\|^2 \leq (1+\varepsilon)L^2\|z_k - z'_k\|^2 + (1+\varepsilon^{-1})\|y_{k+1} - y'_{k+1}\|^2 \tag{5.13}$$

for all $\varepsilon > 0$. In particular (5.13) implies the bound

$$\|z_k - z'_k\|^2 \leq \big((1+\varepsilon)L^2\big)^k\|z_0 - z'_0\|^2 + (1+\varepsilon^{-1})\sum_{j=1}^{k}\big((1+\varepsilon)L^2\big)^{k-j}\|y_j - y'_j\|^2.$$

By summing over k, one notes that φ_n defines a Lipschitz function from $(\mathbb{R}^{nm}, \ell^2)$ into itself, with Lipschitz seminorm

$$L_{\varphi_n} \leq \Big((1+\varepsilon^{-1})\sum_{k=0}^{n-1}\big((1+\varepsilon)L^2\big)^k\Big)^{1/2}$$

We now select $\varepsilon > 0$ according to the value of L : when $L < 1$, we let $\varepsilon = L^{-1} - 1 > 0$, so that $L_{\varphi_n} \leq (1-L)^{-1}$; whereas when $L \geq 1$, we let $\varepsilon = n^{-1}$, and obtain $L_{\varphi_n} \leq [n(n+1)(e-1)]^{(1/2)}$ for $L = 1$, and $L_{\varphi_n} \leq [e(n+1)L^n(L-1)^{-1}]^{1/2}$ for $L > 1$.

Moreover, φ_n induces the joint distribution of $(Z_j)_{j=0}^{n-1}$ from the joint distribution of $(Z_0, Y_1, \dots, Y_{n-1})$. By independence, the joint distribution of $(Z_0, Y_1, \dots, Y_{n-1})$ is a product measure on $(\mathbb{R}^{nm}, \ell^2)$ that satisfies $LSI(\alpha)$. Hence the joint distribution of $(Z_j)_{j=0}^{n-1}$ satisfies $LSI(\alpha)$, where $\alpha = L_{\varphi_n}^{-2}\,\alpha$. □

The linear case gives the following result for ARMA processes.

Proposition 5.8. *Let A and B be $m \times m$ matrices such that the spectral radius ρ of A satisfies $\rho < 1$. Let also Z_0 and Y_j $(j = 1, 2, \dots)$ be mutually independent standard Gaussian $N(0, I_m)$ random variables in $\mathbb{R}^m$. Then, for any $n \geq 1$, the joint distribution of the ARMA process $(Z_j)_{j=0}^{n-1}$, defined by the recurrence relation*

$$Z_{j+1} = AZ_j + BY_{j+1} \qquad (j = 0, 1, \dots),$$

satisfies $LSI(\alpha)$ where

$$\alpha = \Big(\frac{(1-\sqrt{\rho})}{\max\{1, \|B\|\}}\Big)^2\Big(\sum_{j=0}^{\infty}\rho^{-j}\|A^j\|^2\Big)^{-2}.$$

Proof. By Rota's Theorem [89], A is similar to a strict contraction on $(\mathbb{R}^m, \ell^2)$; that is, there exists an invertible $m \times m$ matrix S and a matrix C such that $\|C\| \leq 1$ and $A = \sqrt{\rho} S^{-1} C S$; one can choose the similarity so that the operator norms satisfy

$$\|S\| \|S^{-1}\| \leq \sum_{j=0}^{\infty} \rho^{-j} \|A^j\|^2 < \infty.$$

Hence the ARMA process reduces to the solution of the recurrence relation

$$SZ_{j+1} = \sqrt{\rho}\, CSZ_j + SBY_{j+1} \qquad (j = 0, 1, \dots) \tag{5.14}$$

which involves the $\sqrt{\rho}$-Lipschitz linear map $\Theta : \mathbb{R}^m \to \mathbb{R}^m : \Theta(w) = \sqrt{\rho}\, C\, w$. Given $n \geq 1$, the linear map $\Phi_n : \mathbb{R}^{nm} \to \mathbb{R}^{nm}$, defined to solve (5.14) by

$$(z_0, y_1, \dots, y_n) \mapsto (Sz_0, SBy_1, \dots, SBy_{n-1}) \mapsto (Sz_0, Sz_1, \dots, Sz_{n-1}) \mapsto (z_0, z_1, \dots, z_{n-1}),$$

has operator norm

$$\|\Phi_n\| \leq \|S\| \|S^{-1}\| (1 - \sqrt{\rho})^{-1} \max\{1, \|B\|\};$$

moreover, Φ_n induces the joint distribution of $(Z_0, \dots, Z_{n-1})$ from the joint distribution of $(Z_0, Y_1, \dots, Y_{n-1})$. By Gross's Theorem (see [58]), the latter distribution satisfies $LSI(1)$, and hence the induced distribution satisfies $LSI(\alpha)$, with $\alpha = \|\Phi_n\|^{-2}$. □

Remarks 5.9. (i) As compared to Proposition 5.7, the condition imposed in Proposition 5.8 involves the spectral radius of the matrix A and not its operator norm. In particular, for matrices with norm 1, Proposition 5.7 only leads to LSI with constant of order n^{-2} ; whereas Proposition 5.8 ensures LSI with constant independent of n under the spectral radius assumption $\rho < 1$.

(ii) The joint distribution of the ARMA process is discussed by Djellout, Guillin and Wu [46, Section 3]. We have improved upon [46] by obtaining $LSI(\alpha)$, hence $T_2(\alpha)$, under the spectral radius condition $\rho < 1$, where α is independent of the size n of the considered sample and the size of the matrices.

5.4 Logarithmic Sobolev inequalities for weakly dependent processes

In this section we consider a stochastic process $(\xi_j)_{j=1}^n$, with state space $\mathbb{R}^m$ and initial distribution $P^{(1)}$, which is not necessarily Markovian ; we also assume that the transition kernels have positive densities with respect to Lebesgue measure, and write

$$dp_j = p_j(dx_j \mid x^{(j-1)}) = e^{-u_j(x^{(j)})} dx_j \qquad (j = 2, \dots, n).$$

The coupling between variables is measured by the following integral

$$\Lambda_{j,k}(s) = \sup_{x^{(j-1)}} \int_{\mathbb{R}^m} \exp\Big(\Big\langle s, \frac{\partial u_j}{\partial x_k}(x^{(j)}) \Big\rangle\Big)\, p_j(dx_j \mid x^{(j-1)}), \qquad (s \in \mathbb{R}^m,\ 1 \leq k < j \leq n)$$

where as above $\partial/\partial x_k$ denotes the gradient with respect to $x_k \in \mathbb{R}^m$. The main result in this section is the following.

Theorem 5.10. *Suppose that there exist constants $\alpha > 0$ and $\kappa_{j,k} \geq 0$ for $1 \leq k < j \leq n$ such that*

(i) $P^{(1)}$ and $p_k(. \mid x^{(k-1)})$ ($k = 2, \ldots, n$; $x^{(k-1)} \in \mathbb{R}^{(k-1)m}$) satisfy $LSI(\alpha)$;

(ii) $\Lambda_{j,k}(s) \leq \exp(\kappa_{j,k} \|s\|^2/2)$ holds for all $s \in \mathbb{R}^m$.

Then the joint distribution $P^{(n)}$ satisfies $LSI(\alpha_n)$ with

$$\alpha_n = \frac{\alpha}{1+\varepsilon}\Big(1 + \sum_{k=0}^{n-2} \prod_{j=k+1}^{n-1} (1 + K_j)\Big)^{-1} \tag{5.15}$$

for all $\varepsilon > 0$, where $K_j = (1 + \varepsilon^{-1}) \sum_{\ell=0}^{j-1} \kappa_{n-\ell,n-j}/\alpha$ for $j = 1, \ldots, n-1$.

Suppose further that there exist $R \geq 0$ and $\rho_\ell \geq 0$ for $\ell = 1, \ldots, n-1$ such that

(iii) $\kappa_{j,k} \leq \rho_{j-k}$ for $1 \leq k < j \leq n$, and $\sum_{\ell=1}^{n-1} \sqrt{\rho_\ell} \leq \sqrt{R}$.

Then $P^{(n)}$ satisfies $LSI(\alpha_n)$ where

$$\alpha_n = \begin{cases} \big(\sqrt{\alpha} - \sqrt{R}\big)^2 & \text{if } \ R < \alpha, \\ \dfrac{\alpha}{n(n+1)(e-1)} & \text{if } \ R = \alpha, \\ \Big(\dfrac{\alpha}{R}\Big)^n \dfrac{R-\alpha}{e(n+1)} & \text{if } \ R > \alpha. \end{cases}$$

Before proving this theorem, we give simple sufficient conditions for hypothesis (ii) to hold. When $d = 1$, hypothesis (i) is equivalent to a condition on the cumulative distribution functions by the criterion for LSI given in [17].

Proposition 5.11. *In the above notation, let $1 \leq k < j$ and suppose that there exist $\alpha > 0$ and $L_{j,k} \geq 0$ such that*

(i) $p_j(. \mid x^{(j-1)})$ satisfies $GC(1/\alpha)$ for all $x^{(j-1)} \in \mathbb{R}^{(j-1)m}$;

(ii) u_j is twice continuously differentiable and the off-diagonal blocks of its Hessian matrix satisfy

$$\Big\|\frac{\partial^2 u_j}{\partial x_j \partial x_k}\Big\| \leq L_{j,k}$$

as matrices $(\mathbb{R}^m, \ell^2) \to (\mathbb{R}^m, \ell^2)$.

Then

$$\Lambda_{j,k}(s) \leq \exp(L_{j,k}^2 \|s\|^2/(2\alpha)) \qquad (s \in \mathbb{R}^m).$$

Proof of Proposition 5.11. Letting $s = \|s\|\, e$ for some unit vector e, we note that by (ii) the real function $x_j \mapsto \langle e, \partial u_j/\partial x_k \rangle$ is $L_{j,k}$-Lipschitz in the variable of integration, and that

$$\int_{\mathbb{R}^m} \langle e, \frac{\partial u_j}{\partial x_k} \rangle\, p_j(dx_j \mid x^{(j-1)}) = -\Big\langle e, \frac{\partial}{\partial x_k} \int_{\mathbb{R}^m} p_j(dx_j \mid x^{(j-1)})\Big\rangle = 0$$

since $p_j(. \mid x^{(j-1)})$ is a probability measure. Then, by (i),

$$\int_{\mathbb{R}^m} \exp\Big(\langle s, \frac{\partial u_j}{\partial x_k}\rangle\Big)\, p_j(dx_j \mid x^{(j-1)}) \le \exp(\kappa\, L_{j,k}^2 \,\|s\|^2/2)$$

holds for all $x^{(j-1)}$ in $\mathbb{R}^{(j-1)m}$. This inequality implies the Proposition. □

Proof of Theorem 5.15. For notational convenience, X denotes the state space $\mathbb{R}^m$. Then let $f : X^n \to \mathbb{R}$ be a smooth and compactly supported function, and let $g_j : X^{n-j} \to \mathbb{R}$ be defined by $g_0 = f$ and by

$$g_j(x^{(n-j)}) = \Big(\int_X g_{j-1}(x^{(n-j+1)})^2\, p_{n-j}(dx_{n-j+1} \mid x^{(n-j)})\Big)^{1/2} \tag{5.16}$$

for $j = 1, \ldots, n-1$; finally, let g_n be the constant $(\int_X f^2\, dP^{(n)})^{1/2}$.

From the recursive formula (5.16) one can easily verify the identity

$$\int_{X^n} f^2 \log\Big(f^2/\int_{X^n} f^2 dP^{(n)}\Big) dP^{(n)} = \sum_{j=0}^{n-1} \int_{X^{n-j}} g_j^2 \log\big(g_j^2/g_{j+1}^2\big) dP^{(n-j)} \tag{5.17}$$

which is crucial to the proof ; indeed, it allows us to obtain the result from logarithmic Sobolev inequalities on X.

By hypothesis (i), the measure $dp_{n-j} = p_{n-j}(dx_{n-j} \mid x^{(n-j-1)})$ satisfies $LSI(\alpha)$, whence

$$\int_X g_j^2 \log\big(g_j^2/g_{j+1}^2\big) dp_{n-j} \le \frac{2}{\alpha} \int_X \Big(\frac{\partial g_j}{\partial x_{n-j}}\Big)^2 dp_{n-j} \qquad (j = 0, \ldots, n-1), \tag{5.18}$$

where for $j = n-1$ we take $dp_1 = P^{(1)}(dx_1)$. The next step is to express these derivatives in terms of the gradient of f, using the identity

$$g_j \frac{\partial g_j}{\partial x_{n-j}} = \int_{X^{n-j}} f \frac{\partial f}{\partial x_{n-j}}\, dp_n \ldots dp_{n-j+1} - \frac{1}{2} \sum_{\ell=0}^{j-1} \int_{X^{j-\ell}} g_\ell^2\, \frac{\partial u_{n-\ell}}{\partial x_{n-j}}\, dp_{n-\ell} \ldots dp_{n-j+1} \tag{5.19}$$

which follows from the definition (5.16) of g_j^2 and that of p_{n-j}. The integrals on the right-hand side of (5.19) will be bounded by the following Lemma.

Lemma 5.12. *Let $0 \le \ell < j \le n-1$, and assume that hypothesis (ii) holds. Then*

$$\Big\|\int_X g_\ell^2\, \frac{\partial u_{n-\ell}}{\partial x_{n-j}}\, dp_{n-\ell}\Big\| \le g_{\ell+1} \Big(2\, \kappa_{n-\ell,n-j} \int_X g_\ell^2 \log(g_\ell^2/g_{\ell+1}^2)\, dp_{n-\ell}\Big)^{1/2}. \tag{5.20}$$

Proof of Lemma 5.12. By definition of $\Lambda_{n-\ell,n-j}$, we have

$$\int_X \exp\Big(\langle s, \frac{\partial u_{n-\ell}}{\partial x_{n-j}}\rangle - \log \Lambda_{n-\ell,n-j}(s)\Big) dp_{n-\ell} \le 1 \qquad (s \in X),$$

and hence by the dual formula for relative entropy, as in [16, p. 693],

$$\int_X \Big(\big\langle s, \frac{\partial u_{n-\ell}}{\partial x_{n-j}}\big\rangle - \log \Lambda_{n-\ell,n-j}(s)\Big) g_\ell^2 \, dp_{n-\ell} \le \int_X g_\ell^2 \log\big(g_\ell^2/g_{\ell+1}^2\big) dp_{n-\ell}.$$

Then hypothesis (ii) of the Theorem ensures that

$$\big\langle s, \int_X \frac{\partial u_{n-\ell}}{\partial x_{n-j}} \, g_\ell^2 \, dp_{n-\ell}\big\rangle \le \frac{\|s\|^2}{2} \, \kappa_{n-\ell,n-j} \, g_{\ell+1}^2 + \int_X g_\ell^2 \log\big(g_\ell^2/g_{\ell+1}^2\big) dp_{n-\ell}$$

and the stated result follows by optimizing this over $s \in \mathbb{R}^m$. □

Conclusion of the Proof of Theorem 5.10. When we integrate (5.20) with respect to $dp_{n-\ell-1} \dots dp_{n-j+1}$, we deduce by the Cauchy–Schwarz inequality that

$$\Big| \int_{\mathbb{R}^{j-\ell}} g_\ell^2 \frac{\partial u_{n-\ell}}{\partial x_{n-j}} dp_{n-\ell} \dots dp_{n-j+1} \Big| \\ \le g_j \Big(2 \, \kappa_{n-\ell,n-j} \int_{\mathbb{R}^{j-\ell}} g_\ell^2 \log(g_\ell^2/g_{\ell+1}^2) dp_{n-\ell} \dots dp_{n-j+1} \Big)^{1/2}.$$

Then, by integrating the square of (5.19) with respect to $dP^{(n-j)}$ and making a further application of the Cauchy–Schwarz inequality, we obtain

$$\int_{X^{n-j}} \Big\| \frac{\partial g_j}{\partial x_{n-j}} \Big\|^2 dP^{(n-j)} \le (1+\varepsilon) \int_{X^n} \Big\| \frac{\partial f}{\partial x_{n-j}} \Big\|^2 dP^{(n)} + \frac{1+\varepsilon^{-1}}{4} \Big\{ \sum_{\ell=0}^{j-1} \Big(2 \, \kappa_{n-\ell,n-j} \, h_\ell \Big)^{1/2} \Big\}^2 \tag{5.21}$$

where $\varepsilon > 0$ is arbitrary and h_ℓ is given by

$$h_\ell = \int_{X^{n-\ell}} g_\ell^2 \log\big(g_\ell^2/g_{\ell+1}^2\big) dP^{(n-\ell)}.$$

From (5.21), which holds true for $j = 1, \dots, n-1$, we first prove the general result given in (5.15). By (5.18) and the Cauchy–Schwarz inequality again, we obtain from (5.21) the crucial inequality

$$h_j \le d_j + K_j \sum_{m=0}^{j-1} h_m \qquad (j = 1, \dots, n-1)$$

where we have let

$$\begin{aligned} d_j &= \frac{2(1+\varepsilon)}{\alpha} \int_{\mathbb{R}^n} \Big(\frac{\partial f}{\partial x_{n-j}} \Big)^2 dP^{(n)} \qquad (j = 0, \dots, n-1), \\ K_j &= \frac{1+\varepsilon^{-1}}{\alpha} \sum_{\ell=0}^{j-1} \kappa_{n-\ell,n-j} \qquad (j = 1, \dots, n-1). \end{aligned}$$

Since $h_0 \le d_0$ and all terms are positive, the partial sums $H_k = \sum_{j=0}^k h_j$ satisfy the system of inequalities

$$H_k \le d_k + (1+K_k) H_{k-1} \qquad (k = 1, \dots, n-1),$$

with $H_0 \leq d_0$. By induction, one can deduce that

$$H_{n-1} \leq d_{n-1} + \sum_{k=0}^{n-2} d_k \prod_{\ell=k+1}^{n-1} (1+K_\ell),$$

which in turn implies the bound

$$H_{n-1} \leq \Big(1 + \sum_{k=0}^{n-2} \prod_{\ell=k+1}^{n-1} (1+K_\ell)\Big) \sum_{j=0}^{n-1} d_j.$$

By (5.17) this is equivalent to the inequality

$$\int_{X^n} f^2 \log\Big(f^2 / \int_{X^n} f^2 dP^{(n)}\Big) dP^{(n)} \leq \frac{2(1+\varepsilon)}{\alpha}\Big(1 + \sum_{k=0}^{n-2} \prod_{\ell=k+1}^{n-1} (1+K_\ell)\Big) \int_{X^n} \|\nabla f\|^2 dP^{(n)}.$$

Since f is arbitrary, this ensures that $P^{(n)}$ satisfies $LSI(\alpha_n)$ with α_n as in (5.15).

(iii) The extra hypothesis (iii) enables us to strengthen the preceding inequalities, so (5.21) leads to the convolution-type inequality

$$h_j \leq d_j + \frac{1+\varepsilon^{-1}}{\alpha}\Big(\sum_{\ell=0}^{j-1} \sqrt{\rho_{j-\ell}}\,\sqrt{h_\ell}\Big)^2$$

for $j = 1, \ldots, n-1$, and $h_0 \leq d_0$ for $j = 0$. By summing over j we obtain

$$\sum_{j=0}^{k} h_j \leq \sum_{j=0}^{k} d_j + \frac{1+\varepsilon^{-1}}{\alpha} \sum_{j=1}^{k}\Big(\sum_{\ell=0}^{j-1} \sqrt{\rho_{j-\ell}}\,\sqrt{h_\ell}\Big)^2,$$

which implies by Young's convolution inequality that

$$\sum_{j=0}^{k} h_j \leq \sum_{j=0}^{k} d_j + \frac{1+\varepsilon^{-1}}{\alpha}\Big(\sum_{\ell=1}^{k} \sqrt{\rho_\ell}\Big)^2 \sum_{\ell=0}^{k-1} h_\ell.$$

Now let $R_j = (\sum_{\ell=1}^{j} \sqrt{\rho_\ell})^2$ and $D_j = \sum_{\ell=0}^{j} d_\ell$; then by induction one can prove that

$$H_k \leq D_k + \sum_{j=0}^{k-1} D_j \prod_{\ell=j+1}^{k} \frac{1+\varepsilon^{-1}}{\alpha} R_\ell$$

for $k = 1, \ldots, n-1$, and hence

$$H_{n-1} \leq \Big(1 + \sum_{j=0}^{n-2} \Big(\frac{1+\varepsilon^{-1}}{\alpha} R\Big)^{n-j-1}\Big) D_{n-1} = \sum_{\ell=0}^{n-1} \Big(\frac{1+\varepsilon^{-1}}{\alpha} R\Big)^{\ell} D_{n-1} \tag{5.22}$$

since $D_j \leq D_{n-1}$ and $R_j \leq R$ by hypothesis (iii). We finally select ε to make the bound (5.22) precise, according to the relative values of R and α.

When $R = 0$, we recover $LSI(\alpha)$ for $P^{(n)}$ as expected, since here $P^{(n)}$ is the tensor product of its marginal distributions, which satisfy $LSI(\alpha)$.

When $0 < R < \alpha$, we choose $\varepsilon = (\sqrt{(\alpha/R)}-1)^{-1} > 0$ so that $(1+\varepsilon^{-1})R/\alpha = \sqrt{(R/\alpha)} < 1$ and hence

$$H_{n-1} \leq D_{n-1} \sum_{\ell=0}^{\infty} (R/\alpha)^{\ell/2} = \frac{D_{n-1}}{1-\sqrt{(R/\alpha)}},$$

which by (5.17) and the definition of H_{n-1} and D_{n-1} implies the inequality

$$\int_{X^n} f^2 \log\Big(f^2 / \int_{X^n} f^2 dP^{(n)}\Big) dP^{(n)} \leq \frac{2}{(\sqrt{\alpha}-\sqrt{R})^2} \int_{X^n} \|\nabla f\|^2 dP^{(n)}.$$

When $R \geq \alpha$, we choose $\varepsilon = n$ in (5.22), obtaining

$$H_{n-1} \leq \frac{2(n+1)}{\alpha} \left(\frac{(1+1/n)^n (R/\alpha)^n - 1}{(1+1/n)(R/\alpha)-1}\right) \int_{X^n} \|\nabla f\|^2 dP^{(n)};$$

as above this leads to the stated result by (5.17). This concludes the proof. $\square$

5.5 Logarithmic Sobolev inequalities for Markov processes

The results of the preceding section simplify considerably when we have an homogeneous Markov process $(\xi_j)_{j=1}^n$ with state space $\mathbb{R}^m$, as we shall now show. Suppose that the transition measure is $p(dy \mid x) = e^{-u(x,y)} dy$ where u is a twice continuously differentiable function such that

$$\Lambda(s \mid x) = \int_{\mathbb{R}^m} \exp\Big(\big\langle s, \frac{\partial u}{\partial x}(x,y)\big\rangle\Big)\, p(dy \mid x) < \infty \qquad (s, x \in \mathbb{R}^m). \tag{5.23}$$

Then Theorem 5.10 has the following consequence.

Corollary 5.13. *Suppose that there exist constants $\kappa \geq 0$ and $\alpha > 0$ such that :*

(i) $P^{(1)}$ and $p(. \mid x)$ $(x \in \mathbb{R}^m)$ satisfy $LSI(\alpha)$;

(ii) $\Lambda(s \mid x) \leq \exp(\kappa \|s\|^2/2)$ holds for all $s, x \in \mathbb{R}^m$.

Then the joint law $P^{(n)}$ of the first n variables satisfies $LSI(\alpha_n)$, where

$$\alpha_n = \begin{cases} \big(\sqrt{\alpha} - \sqrt{\kappa}\big)^2 & \text{if } \kappa < \alpha, \\ \dfrac{\alpha}{n(n+1)(e-1)} & \text{if } \kappa = \alpha, \\ \Big(\dfrac{\alpha}{\kappa}\Big)^n \dfrac{\kappa - \alpha}{e(n+1)} & \text{if } \kappa > \alpha. \end{cases}$$

Proof. In the notation of section 5.4, we have $u_j(x^{(j)}) = u(x_{j-1}, x_j)$, so we can take $\kappa_{j,k} = 0$ for $k = 1, \dots, j-2$, and $\kappa_{j,j-1} = \kappa$ for $j = 2, \dots, n$; hence we can take $\rho_1 = \kappa$ and $\rho_j = 0$ for $j = 2, 3, \dots$. Now we can apply Theorem 5.10 (iii) and obtain the stated result with $R = \kappa$ in the various cases. (In fact (5.21) simplifies considerably for a Markov process, and hence one can obtain an easier direct proof of Corollary 5.13.) □

Proof of Theorem 5.3.

By the mean-value theorem and hypothesis (ii) of Theorem 5.3, the function $y \mapsto \langle e, \partial u/\partial x\rangle$ is L-Lipschitz $(\mathbb{R}^m, \ell^2) \to \mathbb{R}$ for any unit vector e in $\mathbb{R}^m$, and hence $\Lambda(s \mid x) \leq \exp(\|s\|^2 L^2/(2\alpha))$ holds for all $s \in \mathbb{R}^m$ as in Proposition 5.11. Hence we can take $\kappa = L^2/\alpha$ in Corollary 5.13 and deduce Theorem 5.3 with the various values of the constant. □

Remark 5.14. Corollary 5.13 is a natural refinement of Theorems 5.1 and 5.2. Indeed $LSI(\alpha)$ implies $T_s(\alpha)$. Then, in the notation of the mentioned results, suppose that u is a twice continuously differentiable function with bounded second-order partial derivatives. Then, by Proposition 5.5, hypotheses (i) and (ii) of Corollary 5.13 together imply that the map $x \mapsto p(. \mid x)$ is $(\kappa/\alpha)^{1/2}$ Lipschitz as a function $\mathbb{R}^m \to (\mathrm{Prob}_2(\mathbb{R}^m), W_2)$, hence $\mathbb{R}^m \to (\mathrm{Prob}_s(\mathbb{R}^m), W_s)$ as in Theorems 5.1 or 5.2. Similarly Proposition 5.5 ensures that Theorem 5.10 is a refinement of Theorem 5.4 with, for $s = 2$,

$$M \leq M_2/\alpha = \frac{1}{\alpha} \sup_{x^{(j-1)}} \sum_{k=1}^{j-1} \int_{\mathbb{R}^m} \Big\| \frac{\partial u_j}{\partial x_k} \Big\|^2 p_j(dx_j \mid x^{(j-1)}) \leq \frac{1}{\alpha} \sum_{k=1}^{j-1} \kappa_{j,k}.$$

Note also the similarity between the constants in Theorem 5.4 (iii) and Theorem 5.10 (iii) when $s = 2$ and one rescales R suitably. In Example 5.15 we show these constants to be optimal.

Example 5.15. (Ornstein–Uhlenbeck Process) We now show that the constants κ_n of Theorem 5.1 (or Theorem 5.6(iii)) and α_n of Corollary 5.13 have optimal growth in n. For this purpose we consider the real Ornstein–Uhlenbeck process conditioned to start at $x \in \mathbb{R}$, namely the solution to the Itô stochastic differential equation

$$dZ_t^{(x)} = -\rho Z_t^{(x)} dt + dB_t^{(0)}, \qquad (t \geq 0)$$

where $(B_t^{(0)})$ is a real standard Brownian motion starting at 0, and $\rho \in \mathbb{R}$. In financial modelling, OU processes with $\rho < 0$ are used to model stock prices in a rising market (see [51, p. 26] for instance). More precisely we consider the discrete-time Markov process $(\xi_j)_{j=1}^n$ defined by $\xi_j = Z_{j\tau}^{(x)}$ where $\tau > 0$, and test the Gaussian concentration inequality with the 1-Lipschitz function $F_n : (\mathbb{R}^n, \ell^1) \to \mathbb{R}$ defined by $F_n(x^{(n)}) = \sum_{j=1}^n x_j$.

The exponential integral satisfies

$$\int_{\mathbb{R}^n} \exp\big(sF_n(x^{(n)})\big)\, P^{(n)}(dx^{(n)}) = \mathbb{E}\, \exp\big(sF_n(\xi^{(n)})\big) = \mathbb{E}\, \exp\Big(s \sum_{j=1}^n Z_{j\tau}^{(x)}\Big). \qquad (5.24)$$

This sum can be expressed in terms of the increments of the OU process

$$\sum_{j=1}^{n} Z_{j\tau}^{(x)} = \sum_{i=1}^{n} \theta^i Z_0^{(x)} + \sum_{j=0}^{n-1} \sum_{i=0}^{n-j-1} \theta^i \big(Z_{(j+1)\tau}^{(x)} - \theta Z_{j\tau}^{(x)}\big),$$

with $\theta = e^{-\rho\tau}$. Moreover one can integrate the stochastic differential equation and prove that $(Z_{(j+1)\tau}^{(x)} - \theta Z_{j\tau}^{(x)})_{0\leq k\leq n-1}$ are independent random variables each with $N(0,\sigma^2)$ distribution, where $\sigma^2 = (1-\theta^2)/(2\,\rho)$ when $\rho \neq 0$, and $\sigma^2 = \tau$ when $\rho = 0$. Hence the exponential integral (5.24) equals

$$\exp\Big(s\sum_{i=1}^{n}\theta^i x\Big)\prod_{j=0}^{n-1}\mathbb{E}\exp\Big[s\Big(\sum_{i=0}^{n-j-1}\theta^i\Big)\big(Z_{(j+1)\tau}^{(x)} - \theta Z_{j\tau}^{(x)}\big)\Big] = \exp\big(s\,\mathbb{E}\,F_n(\xi^{(n)}) + s^2\kappa_n/2\big)$$

where

$$\kappa_n = \sigma^2 \sum_{j=0}^{n-1}\Big(\sum_{i=0}^{n-j-1}\theta^i\Big)^2. \tag{5.25}$$

However, hypothesis (i) of Theorem 5.1 holds with $L = \theta$, since $P^{(1)}$ with distribution $N(x,\sigma^2)$ and $p(.\mid x)$ with distribution $N(\theta x,\sigma^2)$ satisfy $GC(\kappa_1)$ where $\kappa_1 = \sigma^2$, while hypothesis (ii) is satisfied with

$$W_1(p(.\mid x), p(.\mid x')) = W_1(N(\theta x,\sigma^2), N(\theta x',\sigma^2)) = \theta|x-x'| \qquad (x,x'\in\mathbb{R}). \tag{5.26}$$

Hence the constant $\kappa_n(L)$ given by Theorem 5.1 is exactly the directly computed constant κ_n in (5.25), in each of the cases $L=1$, $L>1$ and $L<1$, corresponding to $\rho = 0$, $\rho < 0$ and $\rho > 0$.

As regards Corollary 5.13, note that the transition probability is given by

$$p(dy\mid x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{(y-\theta x)^2}{2\sigma^2}\Big)dy$$

since $Z_\tau^{(x)}$ is distributed as $\theta x + B_{\sigma^2}^{(0)}$. Hence by direct calculation we have

$$\alpha = \frac{1}{\sigma^2},\quad \kappa = \frac{\theta^2}{\sigma^2},\quad L = \frac{\theta}{\sigma^2};$$

consequently the dependence parameters $(\kappa/\alpha)^{1/2}$ and θ given in (5.26) coincide, as in Remark 5.14(ii).

Further, by considering the function $f(x^{(n)}) = \exp\big(\sum_{j=1}^n \theta^j x_j\big)$, one can prove that the joint law $P^{(n)}$ cannot satisfy a logarithmic Sobolev inequality with α_n greater than some constant multiple of n^{-3} for $\theta = 1$, and $(\alpha/\kappa)^n$ for $\theta > 1$. Thus for $\theta \geq 1$, we recover the order of growth in n of the constants given in Corollary 5.13 ; whereas for $\theta < 1$, the constant given in Corollary 5.13 is independent of n.

The OU process does not satisfy the Doeblin condition D_0, as Rosenblatt observes ; see [95, p. 214].

Troisième partie

Systèmes de particules en interaction

Chapitre 6

Premiers résultats de concentration dans la limite de champ moyen

Ce chapitre correspond à l'article [22] écrit en collaboration avec Arnaud Guillin et Cédric Villani, et accepté pour publication dans Probability Theory and Related Fields.

Dans ce chapitre nous établissons tout d'abord des inégalités de concentration quantitatives pour la mesure empirique de variables indépendantes à valeurs dans $\mathbb{R}^d$*, exprimées en distances de Wasserstein. De ceci et d'un argument de couplage nous déduisons des bornes d'erreur dans un problème d'approximation d'une équation de champ moyen par un système de particules stochastiques en interaction.*

Introduction

Large stochastic particle systems constitute a popular way to perform numerical simulations in many contexts, either because they are used in some physical model (as in e.g. stellar or granular media) or as an approximation of a continuous model (as in e.g. vortex simulation for Euler equation, see [76, Chapter 5] for instance). For such systems one may wish to establish concentration estimates showing that the behavior of the system is sharply stabilized as the number N of particles goes to infinity. It is natural to search for these estimates in the setting of large (or moderate) deviations, since one wishes to make sure that the numerical method has a very small probability to give wrong results. From a physical perspective, concentration estimates may be useful to establish the validity of a continuous approximation such as a mean-field limit.

When one is interested in the asymptotic behavior of just one, or a few observables (such as the mean position...), there are efficient methods, based for instance on concentration of measure theory. As a good example, Malrieu [74] recently applied tools from the fields of

Logarithmic Sobolev inequalities, optimal transportation and concentration of measure, to prove very neat bounds like

$$\sup_{\|\varphi\|_{\mathrm{Lip}}\leq 1} \mathbb{P}\left[\Big|\frac{1}{N}\sum_{i=1}^{N}\varphi(X_t^i) - \int \varphi\, d\mu_t\Big| > \frac{C}{\sqrt{N}} + \varepsilon\right] \leq 2\, e^{-\lambda N \varepsilon^2}. \tag{6.1}$$

Here $(X_t^i)_{1\leq i\leq N}$ stand for the positions of particles (in phase space) at time t, ε is a given error, $\mathbb{P}$ stands for the probability, μ_t is a probability measure governing the limit behavior of the system, C and λ are positive constants depending on the particular system he is considering (a simple instance of McKean-Vlasov model used in particular in the modelling of granular media). Moreover,

$$\|\varphi\|_{\mathrm{Lip}} := \sup_{x\neq y} \frac{|\varphi(x)-\varphi(y)|}{d(x,y)},$$

where d is the distance in phase space (say the Euclidean norm $|\cdot|$ in $\mathbb{R}^d$).

This approach can lead to nice bounds, but has the drawback to be limited to a finite number of observables. Of course, one may apply (6.1) to many functions φ, and obtain something like

$$\mathbb{P}\left[\sum_{k=1}^{\infty}\frac{1}{k^2}\Big|\frac{1}{N}\sum_{i=1}^{N}\varphi_k(X_t^i) - \int \varphi_k\, d\mu_t\Big| > \frac{C}{\sqrt{N}} + \varepsilon\right] \leq C e^{-N\lambda\varepsilon^2}, \tag{6.2}$$

where $(\varphi_k)_{k\in\mathbb{N}}$ is an arbitrarily chosen dense family in the set of all 1-Lipschitz functions converging to 0 at infinity. If we denote by δ_x the Dirac mass at point x, and by

$$\hat{\mu}_t^N := \frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}$$

the empirical measure associated with the system (this is a random probability measure), then estimate (6.2) can be interpreted as a bound on how close $\hat{\mu}_t^N$ is to μ_t. Indeed,

$$d(\mu,\nu) := \sum_{k=1}^{\infty}\frac{1}{k^2}\left|\int \varphi_k\, d(\mu-\nu)\right| \tag{6.3}$$

defines a distance on probability measures, associated with a topology which is at least as strong as the weak convergence of measures (convergence against bounded continuous test functions). However, this point of view is deceiving : for practical purposes, the distance d can hardly be estimated, and in any case (6.2) does not contain more information than (6.1) : it is only useful if one considers a finite number of observables.

Sanov's large deviation principle [43, Theorem 6.2.10] provides a more satisfactory tool to estimate the distance between the empirical measure and its limit. Roughly speaking, it implies, for independent variables X_t^i, an estimate of the form

$$\mathbb{P}\left[\mathrm{dist}(\hat{\mu}_t^N,\mu) \geq \varepsilon\right] \simeq e^{-N\alpha(\varepsilon)} \qquad \text{as } N\to\infty,$$

where

$$\alpha(\varepsilon) := \inf\Big\{H(\nu|\mu);\ \mathrm{dist}(\nu,\mu) \geq \varepsilon\Big\} \tag{6.4}$$

and H is the relative H functional :

$$H(\nu|\mu) = \int \frac{d\nu}{d\mu} \ln \frac{d\nu}{d\mu}\, d\mu$$

(to be interpreted as $+\infty$ if ν is not absolutely continuous with respect to μ). Since H behaves in many ways like a square distance, one can hope that $\alpha(\varepsilon) \geq \mathrm{const.}\, \varepsilon^2$. Here "dist" may be any distance which is continuous with respect to the weak topology, a condition which might cause trouble on a non-compact phase space.

Yet Sanov's theorem is not the final answer either : it is actually asymptotic, and only implies a bound like

$$\limsup \frac{1}{N} \log \mathbb{P}\left[\mathrm{dist}(\hat{\mu}_t^N, \mu) \geq \varepsilon\right] \leq -\alpha(\varepsilon),$$

which, unlike (6.1), does not contain any explicit estimate for a given N. Fortunately, there are known techniques to obtain quantitative upper bounds for such theorems, see in particular [43, Exercise 4.5.5]. Since these techniques are devised for compact phase spaces, a further truncation will be necessary to treat more general situations.

In this paper, we shall show how to combine these ideas with recent results about measure concentration and transportation distances, in order to derive in a systematic way estimates that are explicit, deal with the empirical measure as a whole, apply to non-compact phase spaces, and can be used to study some particle systems arising in practical problems. Typical estimates will be of the form

$$\mathbb{P}\left[\sup_{\|\varphi\|_{\mathrm{Lip}}\leq 1} \Big(\frac{1}{N}\sum_{i=1}^N \varphi(X_t^i) - \int \varphi\, d\mu_t\Big) > \varepsilon\right] \leq C e^{-\lambda N \varepsilon^2}. \tag{6.5}$$

As a price to pay, the constant C in the right-hand side will be much larger than the one in (6.1).

Here is a possible application of (6.5) in a numerical perspective. Suppose your system has a limit invariant measure $\mu_\infty = \lim \mu_t$ as $t \to \infty$, and you wish to numerically plot its density f_∞. For that, you run your particle simulation for a long time $t = T$, and plot, say,

$$\tilde{f}_t(x) := \frac{1}{N}\sum_{i=1}^N \zeta_\alpha\big(x - X_t^i\big), \tag{6.6}$$

where $\zeta_\alpha = \alpha^{-d}\zeta(x/\alpha)$ is a smooth approximation of a Dirac mass as $\alpha \to 0$ (as usual, ζ is a nonnegative smooth radial function on $\mathbb{R}^d$ with compact support and unit integral). With the help of estimates such as (6.5), it is often possible to compute bounds on, say,

$$\mathbb{P}\Big[\|\tilde{f}_T - f_\infty\|_{L^\infty} > \varepsilon\Big]$$

in terms of N, ε, T and α. In this way one can "guarantee" that all details of the invariant measure are captured by the stochastic system. While this problem is too general to be treated abstractly, we shall show on some concrete model examples how to derive such bounds for the same kind of systems that was considered by Malrieu.

In the next section, we shall explain about our main tools and results ; the rest of the paper will be devoted to the proofs. Some auxiliary estimates of general interest are postponed in Appendix.

6.1 Tools and main results

6.1.1 Wasserstein distances

To measure distances between probability measures, we shall use transportation distances, also called **Wasserstein distances**. They can be defined in an abstract Polish space X as follows : given p in $[1,+\infty)$, d a lower semi-continuous distance on X, and μ and ν two Borel probability measures on X, the Wasserstein distance of order p between μ and ν is

$$W_p(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \left(\iint d(x,y)^p d\pi(x,y)\right)^{1/p}$$

where π runs over the set $\Pi(\mu,\nu)$ of all joint probability measures on the product space $X\times X$ with marginals μ and ν ; it is easy to check [111, Theorem 7.3] that W_p is a distance on the set $P_p(X)$ of Borel probability measures μ on X such that $\int d(x_0,x)^p\, d\mu(x) < +\infty$.

For this choice of distance, in view of Sanov's theorem, a very natural class of inequalities is the family of so-called transportation inequalities, or **Talagrand inequalities** (see [68] for instance) : by definition, given $p\geq 1$ and $\lambda>0$, a probability measure μ on X satisfies $T_p(\lambda)$ if the inequality

$$W_p(\nu,\mu) \leq \sqrt{\frac{2}{\lambda}\, H(\nu|\mu)}$$

holds for any probability measure ν. We shall say that μ satisfies a T_p inequality if it satisfies $T_p(\lambda)$ for some $\lambda>0$. By Jensen's inequality, these inequalities become stronger as p becomes larger ; so the weakest of all is T_1. Some variants introduced in [23] will also be considered.

Of course T_p is not a very explicit condition, and a priori it is not clear how to check that a given probability measure satisfies it. It has been proven [17, 46, 23] that T_1 is *equivalent* to the existence of a square-exponential moment : in other words, a reference measure μ satisfies T_1 if and only if there is $\alpha>0$ such that

$$\int e^{\alpha d(x,y)^2}\, d\mu(x) < +\infty$$

for some (and thus any) $y\in X$. If that condition is satisfied, then one can find explicitly some λ such that $T_1(\lambda)$ holds true : see for instance [23].

This criterion makes T_1 a rather convenient inequality to use. Another popular inequality is T_2, which appears naturally in many situations where a lot of structure is available, and which has good tensorization properties in many dimensions. Up to now, T_2 inequalities have not been so well characterized : it is known that they are implied by a Logarithmic Sobolev inequality [88, 16, 113], and that they imply a Poincaré, or spectral gap, inequality [88, 16]. See [39] for an attempt to a criterion for T_2. In any case, contrary to the case $p = 1$, there is no hope to obtain T_2 inequalities from just integrability or decay estimates.

In this paper, we shall mainly focus on the case $p = 1$, which is much more flexible.

6.1.2 Metric entropy

When X is a compact space, the minimum number $m(X, r)$ of balls of radius r needed to cover X is called the **metric entropy** of X. This quantity plays an important role in quantitative variants of Sanov's Theorem [43, Exercise 4.5.5]. In the present paper, to fix ideas we shall always be working in the particular Euclidean space $\mathbb{R}^d$, which of course is not compact ; and we shall reduce to the compact case by truncating everything to balls of finite radius R. This particular choice will influence the results through the function $m(\mathcal{P}_p(B_R), r)$, where B_R is the ball of radius R centered at some point, say the origin, and $\mathcal{P}_p(B_R)$ is the space of probability measures on B_R, metrized by W_p.

6.1.3 Sanov-type theorems

The core of our estimates is based on variants of Sanov's Theorem, all dealing with *independent* random variables. Let μ be a given probability measure on $\mathbb{R}^d$, and let $(X^i)_{i=1,\dots,N}$ be a sample of independent variables, all distributed according to μ ; let also

$$\hat{\mu}^N := \frac{1}{N} \sum_{i=1}^{N} \delta_{X^i}$$

be the associated empirical measure. In our first main result we assume a T_p inequality for the measure μ, and deduce from that an upper bound in W_p distance :

Theorem 6.1. *Let $p \in [1, 2]$ and let μ be a probability measure on $\mathbb{R}^d$ satisfying a $T_p(\lambda)$ inequality. Then, for any $d' > d$ and $\lambda' < \lambda$, there exists some constant N_0, depending on λ', d' and some square-exponential moment of μ, such that for any $\varepsilon > 0$ and $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$,*

$$\mathbb{P}\left[W_p(\mu, \hat{\mu}^N) > \varepsilon\right] \leq e^{-\gamma_p \frac{\lambda'}{2} N \varepsilon^2}, \tag{6.7}$$

where

$$\gamma_p = \begin{cases} 1 \ \textit{if}\ 1 \leq p < 2 \\ 3 - 2\sqrt{2} \ \textit{if}\ p = 2 \ . \end{cases}$$

Compared to Sanov's Theorem, this result is more restrictive in the sense that it requires some extra assumptions on the reference measure μ, but under these hypotheses we are able to replace a result which was only asymptotic by a pointwise upper bound on the error probability, together with a lower bound on the required size of the sample.

In view of the Kantorovich-Rubinstein duality formula

$$W_1(\mu,\nu) = \sup \left\{ \int f \, d(\mu-\nu);\ \ \|f\|_{\mathrm{Lip}} \leq 1 \right\}, \tag{6.8}$$

Theorem 6.1 implies concentration inequalities such as

$$\mathbb{P}\left[\sup_{f;\, \|f\|_{\mathrm{Lip}}\leq 1} \Bigl(\frac{1}{N} \sum_{k=1}^{N} f(X_i) - \int f \, d\mu \Bigr) > \varepsilon \right] \leq e^{-\frac{\lambda'}{2} N\varepsilon^2} \tag{6.9}$$

for $\lambda' < \lambda$, and N sufficiently large, under the assumption that μ satisfies a T_1 inequality, or equivalently admits a finite square-exponential moment. Those types of inequalities are of interest in non-parametric statistics and choice models [81].

Remark 6.2. The sole inequality $T_1(\lambda)$ implies that for all 1-Lipschitz function f,

$$\mathbb{P}\left[\frac{1}{N} \sum_{k=1}^{N} f(X_i) - \int f \, d\mu > \varepsilon \right] \leq e^{-\frac{\lambda}{2} N\varepsilon^2}, \tag{6.10}$$

and it is easy to see that the coefficient λ in this inequality is the best possible. While the quantity controlled in Theorem 6.1 is much stronger, the estimate is weakened only in that λ is replaced by some $\lambda' > \lambda$ (arbitrarily close to λ) and that N has to be large enough. In fact, a variant of the proof below would yield estimates such as

$$\mathbb{P}\left[W_p(\mu, \hat{\mu}^N) > \varepsilon \right] \leq C(\varepsilon)\, e^{-\gamma \frac{\lambda'}{2} N \varepsilon^2},$$

where now there is no restriction on N, but $C(\varepsilon)$ is a larger constant, explicitly computable from the proof.

Remark 6.3. As pointed out to us by M. Ledoux, there is another way to concentration estimates on the empirical measure when $d = p = 1$. Indeed, in this specific case,

$$W_1(\hat{\mu}^N, \mu) = \left\| \frac{1}{N} \sum_{i=1}^{N} H(\cdot - X_i) - F \right\|_{L^1(\mathbb{R})}$$

where $H = \mathbf{1}_{[0,+\infty)}$ stands for the Heaviside function on $\mathbb{R}$ and F denotes the repartition function of μ, so that

$$\mathbb{P}\Bigl[W_1(\hat{\mu}^N, \mu) \geq \varepsilon \Bigr] = \mathbb{P}\left[\Bigl\| \frac{1}{N} \sum_{i=1}^{N} F_i \Bigr\|_{L^1} > \varepsilon \right]$$

where

$$F_i := H(\cdot - X_i) - F \qquad (1 \leq i \leq N)$$

are centered $L^1(\mathbb{R})$-valued independent identically distributed random variables. But, according to [6, Exercise 3.8.14], a centered $L^1(\mathbb{R})$-valued random variable Y satisfies a Central Limit Theorem if and only if

$$\int_{\mathbb{R}} \left(\mathbb{E}[Y^2(t)]\right)^{1/2} \, dt < +\infty,$$

a condition which for the random variables F_i's can be written

$$\int_{\mathbb{R}} \sqrt{F(t)(1-F(t))} \, dt < +\infty. \tag{6.11}$$

Condition (6.11) in turn holds true as soon as (for instance) $\int_{\mathbb{R}} |x|^{2+\delta} \, d\mu(x)$ is finite for some positive δ. Then we may apply a quantitative version of the Central Limit Theorem for random variables in the Banach space $L^1(\mathbb{R})$. See [55] and [70] for related works.

Remark 6.4. Theorem 6.1 applies if N is at least as large as ε^{-r} for some $r > d+2$; we do not know whether $d+2$ here is optimal although, for $p=1$, a variant of the present proof, consisting in directly deducing (6.9) from (6.10), also leads to a similar condition $N \geq N_0 \, \varepsilon^{-r}$ with $r > d+2$.

For the applications that we shall treat, in which the tails of the probability distributions will be decaying very fast, Theorem 6.1 will be sufficient. However, it is worthwile pointing out that the technique works under much broader assumptions : weaker estimates can be proven for probability measures that do not decay fast enough to admit finite square-exponential moments. Here below are some such results using only polynomial moment estimates :

Theorem 6.5. *Let $q \geq 1$ and let μ be a probability measure on $\mathbb{R}^d$ such that*

$$\int_{\mathbb{R}^d} |x|^q \, d\mu(x) < +\infty.$$

Then

(i) For any $p \in [1, q/2)$, $\delta \in (0, q/p - 2)$ and $d' > d$, there exists a constant N_0 such that

$$\mathbb{P}\left[W_p(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \varepsilon^{-q} N^{-\frac{q}{2p}+\frac{\delta}{2}}$$

for any $\varepsilon > 0$ and $N \geq N_0 \, \max(\varepsilon^{-q\frac{2p+d'}{q-p}}, \varepsilon^{d'-d})$;

(ii) For any $p \in [q/2, q)$, $\delta \in (0, q/p - 1)$ and $d' > d$ there exists a constant N_0 such that

$$\mathbb{P}\left[W_p(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \varepsilon^{-q} N^{1-\frac{q}{p}+\delta}$$

for any $\varepsilon > 0$ and $N \geq N_0 \max(\varepsilon^{-q\frac{2p+d'}{q-p}}, \varepsilon^{d'-d})$.

Here are also some variants under alternative “regularity” assumptions :

Theorem 6.6. *(i) Let $p \geq 1$; assume that $\mathcal{E}_\alpha := \int e^{\alpha|x|} d\mu$ is finite for some $\alpha > 0$. Then, for all $d' > d$, there exist some constants K and N_0, depending only on d, α and $\mathcal{E}_\alpha$, such that*

$$\mathbb{P}\left[W_p(\mu, \hat{\mu}^N) > \varepsilon\right] \leq e^{-K\,N^{1/p}\,\min(\varepsilon,\varepsilon^2)}$$

for any $\varepsilon > 0$ and $N \geq N_0 \max(\varepsilon^{-(2p+d')}, 1)$.

(ii) Suppose that μ satifies T_1 and a Poincaré inequality, then for all $a < 2$ there exists some constants K and N_0 such that

$$\mathbb{P}\left[W_2(\mu, \hat{\mu}^N) > \varepsilon\right] \leq e^{-K\,N\,\min(\varepsilon^2,\varepsilon^a)} \tag{6.12}$$

for any $\varepsilon > 0$ and $N \geq N_0 \max(\varepsilon^{-(4+d')}, 1)$.

(iii) Let $p > 2$ and let μ be a probability measure on $\mathbb{R}^d$ satisfying $T_p(\lambda)$. Then for all $\lambda' < \lambda$ and $d' > d$ there exists some constant N_0, depending on μ only through λ and some square-exponential moment, such that

$$\mathbb{P}\left[W_p(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \min\left(e^{-\frac{\lambda'}{2}N\varepsilon^2} + e^{-(N\varepsilon^{d'+2})^{2/d'}},\, 2\,e^{-\frac{\lambda'}{4}N^{2/p}\varepsilon^2}\right) \tag{6.13}$$

for any $\varepsilon > 0$ and $N \geq N_0\, \max(\varepsilon^{-(d'+2)}, 1)$.

6.1.4 Interacting systems of particles

We now consider a system of N interacting particles whose time-evolution is governed by the system of coupled stochastic differential equations

$$dX_t^i = \sqrt{2}\,dB_t^i - \nabla V(X_t^i)dt - \frac{1}{N}\sum_{j=1}^{N} \nabla W(X_t^i - X_t^j)dt, \qquad i = 1, \ldots, N. \tag{6.14}$$

Here X_t^i is the position at time t of particule number i, the B^i's are N independent Brownian motions, and V and W are smooth potentials, sufficiently nice that (6.14) can be solved globally in time. We shall always assume that W (which can be interpreted as an interaction potential) is a symmetric function, that is $W(-z) = W(z)$ for all $z \in \mathbb{R}^d$.

Equation (6.14) is a particularly simple instance of coupled system ; in the case when V is quadratic and W has cubic growth, it was used as a simple mean-field kinetic model for granular media (see e.g. [74]). While many of our results could be extended to more general systems, that particular one will be quite enough for our exposition.

To this system of particles is naturally associated the empirical measure, defined for each time $t \geq 0$ by

$$\hat{\mu}_t^N := \sum_{i=1}^{N} \delta_{X_t^i}. \tag{6.15}$$

Under suitable assumptions on the potentials V and W, it is a classical result that, if the initial positions of the particle system are distributed chaotically (for instance, if they are identically distributed, independent random variables), then the empirical measure $\hat{\mu}_t^N$ converges as $N \to \infty$ to a solution of the nonlinear partial differential equation

$$\frac{\partial \mu_t}{\partial t} = \Delta \mu_t + \nabla \cdot \Big(\mu_t \nabla \big(V + W * \mu_t \big) \Big), \tag{6.16}$$

where $\nabla \cdot$ stands for the divergence operator. Equation 6.16 is a simple instance of McKean-Vlasov equation. This convergence result is part of the by now well-developed theory of propagation of chaos, and was studied by Sznitman for pedagogical reasons [105], in the case of potentials that grow at most quadratically at infinity. Later, Benachour, Roynette, Talay and Vallois [9, 10] considered the case where the interaction potential grows faster than quadratically. As far as the limit equation (6.16) is concerned, a discussion of its use in the modelling of granular media in kinetic theory was performed by Benedetto, Caglioti, Carrillo and Pulvirenti [11, 12], while the asymptotic behavior in large time was studied by Carrillo, McCann and Villani [36, 37] with the help of Wasserstein distances and entropy inequality methods. Then Malrieu [74] presented a detailed study of both limits $t \to \infty$ and $N \to \infty$ by probabilistic methods, and established estimates of the type of (6.1) under adequate convexity assumptions on V and W (see also [111, Problem 15]).

As announced before, we shall now give some estimates on the convergence at the level of the law itself. To fix ideas, we assume that V and W have locally bounded Hessian matrices satisfying

$$\begin{cases} \text{(i)} \quad D^2V(x) \geq \beta I, \quad \gamma I \leq D^2 W(x) \leq \gamma' I, \qquad \forall x \in \mathbb{R}^d, \\ \\ \text{(ii)} \quad |\nabla V(x)| = O(e^{a|x|^2}) \quad \text{for any } a > 0. \end{cases} \tag{6.17}$$

Under these assumptions, we shall derive the following bounds.

Theorem 6.7. *Let μ_0 be a probability measure on $\mathbb{R}^d$, admitting a finite square-exponential moment :*

$$\exists \alpha_0 > 0; \qquad M_{\alpha_0} := \int e^{\alpha_0 |x|^2} \, d\mu_0(x) < +\infty.$$

Let $(X_0^i)_{1 \leq i \leq N}$ be N independent random variables with common law μ_0. Let (X_t^i) be the solution of (6.14) with initial value $(X_0^1, \dots X_0^N)$, where V and W are assumed to satisfy (6.17) ; and let μ_t be the solution of (6.16) with initial value μ_0. Let also $\hat{\mu}_t^N$ be the empirical measure associated with the $(X_t^i)_{1 \leq i \leq N}$. Then, for all $T \geq 0$, there exists some constant $K = K(T)$ such that, for any $d' > d$, there exists some constants N_0 and C such that for all $\varepsilon > 0$

$$N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1) \Longrightarrow \quad \mathbb{P} \left[\sup_{0 \leq t \leq T} W_1(\hat{\mu}_t^N, \mu_t) > \varepsilon \right] \leq C(1 + T\varepsilon^{-2}) \exp\big(-K\, N\, \varepsilon^2\big) .$$

Note that in the above theorem we have proven not only that for all t, the empirical measure is close to the limit measure, but also that the probability of observing any significant deviation during a whole time period $[0, T]$ is small.

The fact that $\hat{\mu}_t^N$ is very close to the deterministic measure μ_t implies the propagation of chaos : two particles drawn from the system behave independently of each other as $N \to \infty$ (see Sznitman [105] for more details). But we can also directly study correlations between particles and find more precise estimates : for that purpose it is convenient to consider the empirical measure on *pairs* of particles, defined as

$$\hat{\mu}_t^{N,2} := \frac{1}{N(N-1)} \sum_{i \neq j} \delta_{(X_t^i, X_t^j)}.$$

By a simple adaptation of the computations appearing in the proof of Theorem 6.7, one can prove

Theorem 6.8. *With the same notation and assumptions as in Theorem 6.7, for all $T \geq 0$ and $d' > d$, there exists some constants $K > 0$ and N_0 such that for all $\varepsilon > 0$*

$$N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1) \Longrightarrow \quad \mathbb{P}\left[\sup_{0 \leq t \leq T} W_1(\hat{\mu}_t^{N,2}, \mu_t \otimes \mu_t) > \varepsilon\right] \leq \exp\left(-K\, N\, \varepsilon^2\right).$$

(Here W_1 stands for the Wasserstein distance or order 1 on $P_1(\mathbb{R}^d \times \mathbb{R}^d)$.) Of course, one may similarly consider the problem of drawing k particles with $k \geq 2$.

Theorems 6.7 and 6.8 use Theorem 6.1 as a crucial ingredient, which is why a strong integrability assumption is imposed on μ_0. Note however that, under stronger assumptions on the behaviour at infinity of V or W, as the existence of some $\beta \in \mathbb{R}$, B, $\varepsilon > 0$ such as

$$D^2 V(x) \geq (B|x|^\varepsilon + \beta) I, \qquad \forall x \in \mathbb{R}^d,$$

it can be proven that any square exponential moment for μ_t becomes instantaneously finite for $t > 0$. Note also that, by using Theorem 6.5, one can obtain weaker but still relevant results of concentration of the empirical measure under just polynomial moment assumptions on μ_0, provided that ∇V does not grow too fast at infinity. To limit the size of this paper, we shall not go further into such considerations.

6.1.5 Uniform in time estimates

In the "uniformly convex case" when $\beta > 0, \beta + 2\gamma > 0$, it can be proven [36, 37, 74] that μ_t converges exponentially fast, as $t \to \infty$, to some equilibrium measure μ_∞. In that case, it is natural to expect that the empirical measure is a good approximation of μ_∞ as $N \to \infty$ and $t \to \infty$, uniformly in time. This is what we shall indeed prove :

Theorem 6.9. *With the same notation and assumptions as in Theorem 6.7, suppose that $\beta > 0, \beta + 2\gamma > 0$. Then there exists some constant $K > 0$ such that for any $d' > d$, there exists some constants C and N_0 such that for all $\varepsilon > 0$*

$$N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1) \Longrightarrow \quad \sup_{t \geq 0} \mathbb{P}\left[W_1(\hat{\mu}_t^N, \mu_t) > \varepsilon\right] \leq C(1 + \varepsilon^{-2})\, \exp\left(-K\, N\, \varepsilon^2\right)$$

As a consequence, there are constants T_0, ε_0 (depending on the initial datum) and $K' = K/4$ such that, under the same conditions on N and ε,

$$\sup_{t \geq T_0 \log(\varepsilon_0/\varepsilon)} \mathbb{P}\left[W_1(\hat{\mu}_t^N, \mu_\infty) > \varepsilon\right] \leq C(1 + \varepsilon^{-2})\, \exp\left(-K'\, N\, \varepsilon^2\right).$$

Remark 6.10. In view of the results in [36], it is natural to expect that a similar conclusion holds true when $V = 0$ and W is convex enough. Propositions 6.14 and 6.21 below extend to that case, but it seems trickier to adapt the proof of Proposition 6.23.

We conclude with an application to the numerical reconstruction of the invariant measure.

Theorem 6.11. *With the same notation and assumptions as in Theorem 6.9, consider the mollified empirical measure* (6.6)*. Then one can choose $\alpha = O(\varepsilon)$ in such a way that*

$$N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1) \Longrightarrow \sup_{t \geq T_0 \log(\varepsilon_0/\varepsilon)} \mathbb{P}\left[\|\tilde{f}_t - f_\infty\|_{L^\infty} > \varepsilon\right] \leq C(1+\varepsilon^{-(2d+4)}) \, \exp\left(-K' \, N \, \varepsilon^{2d+4}\right).$$

These results are effective : all the constants therein can be estimated explicitly in terms of the data.

6.1.6 Strategy and plan

The strategy is rather systematic. First, we shall establish Sanov-type bounds for independent variables in $\mathbb{R}^d$ (not depending on time), resulting in concentration results such as Theorems 6.1 to 6.6. This will be achieved along the ideas in [43, Exercices 4.5.5 and 6.2.19] (see also [98, Section 5]), by first truncating to a compact ball, and then covering the set of probability measures on this ball by a finite number of small balls (in the space of probability measures) ; the most tricky part will actually lie in the optimization of parameters.

With such results in hand, we will start the study of the particle system by introducing the nonlinear partial differential equation (6.16). For this equation, the Cauchy problem can be solved in a satisfactory way, in particular existence and uniqueness of a solution, which for $t > 0$ is reasonably smooth, can be shown under various assumptions on V and W (see e.g. [36, 37]). Other regularity estimates such as the decay at infinity, or the smoothness in time, can be established ; also the convergence to equilibrium in large time can sometimes be proven.

Next, following the presentation by Sznitman [105], we introduce a family of independent processes $(Y_t^i)_{1 \leq i \leq N}$, governed by the stochastic differential equation

$$\left\{ \begin{array}{rcl} dY_t^i &=& \sqrt{2}\, dB_t^i - \nabla V(Y_t^i)\, dt - \nabla W * \mu_t(Y_t^i)\, dt, \\ Y_0^i &=& X_0^i. \end{array} \right. \tag{6.18}$$

As a consequence of Itô's formula, the law ν_t of each Y_t^i is a solution of the linear partial differential equation

$$\frac{\partial \nu_t}{\partial t} = \Delta \nu_t + \nabla \cdot \Big(\nabla \big(V + W * \mu_t\big) \nu_t \Big), \qquad \nu_0 = \mu_0.$$

But this linear equation is also solved by μ_t, and a uniqueness theorem implies that actually $\nu_t = \mu_t$, for all $t \geq 0$. See [9, 10] for related questions on the stochastic differential equation (6.18).

For each given t, the independence of the variables Y_t^i and the good decay of μ_t will imply a strong concentration of the empirical measure

$$\hat{\nu}_t^N := \frac{1}{N}\sum_{i=1}^N \delta_{Y_t^i}.$$

To go further, we shall establish a more precise information, such as a control on

$$\mathbb{P}\left[\sup_{0\le t\le T} W_1(\hat{\nu}_t^N, \mu_t) > \varepsilon\right].$$

Such bounds will be obtained by combining the estimate of concentration at fixed time t with some estimates of regularity of $\hat{\nu}_t^N$ (and μ_t) in t, obtained via basic tools of stochastic differential calculus (in particular Doob's inequality).

Finally, we can show by a Gronwall-type argument that the control of the distance of $\hat{\mu}_t^N$ to μ_t reduces to the control of the distance of $\hat{\nu}_t^N$ to μ_t : for instance,

$$\mathbb{P}\left[\sup_{0\le t\le T} W_1(\hat{\mu}_t^N, \mu_t) > \varepsilon\right] \le \mathbb{P}\left[\sup_{0\le t\le T} W_1(\hat{\nu}_t^N, \mu_t) > C\varepsilon\right] \tag{6.19}$$

for some constant C. We shall also show how a variant of this computation provides estimates of the type of those in Theorem 6.9, and how to get data reconstruction estimates as in Theorem 6.11.

6.1.7 Remarks and further developments

The results in this paper confirm what seems to be a rather general rule about Wasserstein distances : results in distance W_1 are very robust and can be used in rather hard problems, with no particular structure ; on the contrary, results in distance W_2 are stronger, but usually require much more structure and/or assumptions. For instance, in the study of the equation (6.16), the distance W_2 works beautifully, and this might be explained by the fact that (6.16) has the structure of a gradient flow with respect to the W_2 distance [36, 37]. In the problem considered by Malrieu [74], W_2 is also well-adapted, but leads him to impose strong assumptions on the initial datum μ_0, such as the existence of a Logarithmic Sobolev inequality for μ_0, considered as a reference measure. As a general rule, in a context of geometric inequalities with more or less subtle isoperimetric content, related to Brenier's transportation mapping theorem, W_2 is also the most natural distance to use [111]. On the contrary, here we are considering quite a rough problem (concentration for the law of a random probability measure, driven by a stochastic differential equation with coupling) and we wish to impose only natural integrability conditions ; then the distance W_1 is much more convenient.

Further developments could be considered. For instance, one may desire to prove some deviation inequalities for dependent sequences, say Markov chains, as both Sanov's theorem and transportation inequality can be established under appropriate ergodicity and integrability conditions.

Considering again the problem of the particle system, in a numerical context, one may wish to take into account the numerical errors associated with the time-discretization of the dynamics (say an implicit Euler scheme). For concentration estimates in one observable, a beautiful study of these issues was performed by Malrieu [75]. For concentration estimates on the whole empirical measure, to our knowledge the study remains to be done. Also errors due to the boundedness of the phase space actually used in the simulation might be taken into account, etc.

At a more technical level, it would be desirable to relax the assumption of boundedness of D^2W in Theorem 6.7, so as to allow for instance the interesting case of cubic interaction. This is much more technical and will be considered in a separate work.

Another issue of interest would be to consider concentration of the empirical measure *on path space*, i.e.

$$\hat{\mu}^N_{[0,T]} := \frac{1}{N}\sum_{i=1}^N \delta_{(X^i_t)_{0\le t\le T}},$$

where T is a fixed time length. Here $\hat{\mu}^N_{[0,T]}$ is a random measure on $C([0,T];\mathbb{R}^d)$ and we would like to show that it is close to the law of the trajectories of the nonlinear stochastic differential equation

$$dY_t = \sqrt{2}\,dB_t - \nabla V(Y_t)\,dt - (\nabla W * \mu_t)(Y_t)\,dt, \tag{6.20}$$

where the initial datum Y_0 is drawn randomly according to μ_0. This will imply a quantitative information on the whole trajectory of a given particle in the system.

When one wishes to adapt the general method to this question, a problem immediately occurs : not only is $C([0,T];\mathbb{R}^d)$ not compact, but also balls with finite radius in this space are not compact either (of course, this is true even if the phase space of particles is compact). One may remedy to this problem by embedding $C([0,T];B_R)$ into a space such as $L^2([0,T];B_R)$, equipped with the weak topology ; but we do not know of any "natural" metric on that space. There is (at least) another way out : we know from classical stochastic processes theory that integral trajectories of differential equations driven by white noise are typically Hölder-α for any $\alpha < 1/2$. This suggests a natural strategy : choose any fixed $\alpha \in (0,1/2)$ and work in the space $\mathcal{C}^\alpha([0,T];\mathbb{R}^d)$, equipped with the norm

$$\|w\|_\alpha := \sup_{0\le t\le T}|w(t)| + \sup_{s\neq t}\frac{|w(t)-w(s)|}{|t-s|^\alpha}.$$

For any $R > 0$, the ball of radius R and center 0 (the zero function) in $\mathcal{C}^\alpha$ is compact, and one may estimate its metric entropy. Then one can hope to perform all estimates by using the norm $\mathcal{C}^\alpha$; for instance, establish a bound on, say, a square-exponential moment on the law of Y_t :

$$\mathbb{E}\exp\left(\beta\|(Y_t)_{0\le t\le T}\|^2_\alpha\right) < +\infty.$$

Again, to avoid expanding the size of the present paper too much, these issues will be addressed separately.

6.2 The case of independent variables

In this section we consider the case where we are given N *independent* variables $X^i \in \mathbb{R}^d$, distributed according to a certain law μ. There is no time dependence at this stage. We shall first examine the case when the law μ has very fast decay (Theorem 6.1), then variants in which it decays in a slower way (Theorem 6.5 and 6.6).

6.2.1 Proof of Theorem 6.1

The proof splits into three steps : (1) Truncation to a compact ball B_R of radius R, (2) covering of $\mathcal{P}(B_R)$ by small balls of radius r and Sanov's argument, and (3) optimization of the parameters.

Step 1 : Truncation. Let $R > 0$, to be chosen later on, and let B_R stand for the ball of radius R and center 0 (say) in $\mathbb{R}^d$. Let $\mathbf{1}_{B_R}$ stand for the indicator function of B_R. We truncate μ into a probability measure μ_R on the ball B_R :

$$\mu_R = \frac{\mathbf{1}_{B_R}\, \mu}{\mu[B_R]}.$$

We wish to bound the quantity $\mathbb{P}\left[W_p(\hat{\mu}^N, \mu) > \varepsilon\right]$ in terms of μ_R and the associated empirical measure. For this purpose, consider independent variables $(X^k)_{1\leq k\leq N}$ drawn according to μ, and $(Y^k)_{1\leq k\leq N}$ drawn according to μ_R, independent of each other ; then define

$$X_R^k := \begin{cases} X^k & \text{if } |X^k| \leq R \\ Y^k & \text{if } |X^k| > R. \end{cases}$$

Since X^1 and X_R^1 are distributed according to μ and μ_R respectively, we have, by definition of Wasserstein distance,

$$\begin{aligned} W_p^p(\mu, \mu_R) \leq \mathbb{E}|X^1 - X_R^1|^p = \mathbb{E}\Big(|X^1 - Y^1|^p \mathbf{1}_{|X^1|>R}\Big) &\leq 2^p \mathbb{E}\big(|X^1|^p \mathbf{1}_{|X^1|>R}\big) \\ &= 2^p \int_{\{|x|>R\}} |x|^p \, d\mu(x). \end{aligned}$$

But μ satisfies a $T_p(\lambda)$ inequality for some $p \geq 1$, hence a fortiori a $T_1(\lambda)$ inequality, so

$$E_\alpha := \int_{\mathbb{R}^d} e^{\alpha|x|^2} \, d\mu(x) < +\infty$$

for some $\alpha > 0$ (any $\alpha < \lambda/2$ would do). If R is large enough (say, $R \geq \sqrt{p/(2\alpha)}$), then the function $r \longmapsto \dfrac{r^p}{e^{\alpha r^2}}$ is nonincreasing for $r \geq R$, and then

$$W_p^p(\mu, \mu_R) \leq 2^p \left(\frac{R^p}{e^{\alpha R^2}}\right) \int_{\{|x|>R\}} e^{\alpha|x|^2} \, d\mu(x).$$

We conclude that

$$W_p^p(\mu, \mu_R) \leq 2^p E_\alpha R^p e^{-\alpha R^2} \qquad (\alpha < \lambda/2,\ R \geq \sqrt{p/2\alpha}). \tag{6.21}$$

On the other hand, the empirical measures

$$\hat{\mu}^N := \frac{1}{N}\sum_{k=1}^N \delta_{X^k}, \qquad \hat{\mu}_R^N := \frac{1}{N}\sum_{k=1}^N \delta_{X_R^k}$$

satisfy

$$W_p^p(\hat{\mu}_R^N, \hat{\mu}^N) \leq \frac{1}{N}\sum_{k=1}^N |X_R^k - X^k|^p \leq \frac{1}{N}\sum_{k=1}^N Z^k,$$

where $Z^k := 2^p\,|X^k|^p\,\mathbf{1}_{|X^k|>R}$ $(k = 1,\ldots,N)$. Then, for any $p \in [1,2]$, we can introduce parameters ε and $\theta > 0$, and use Chebyshev's exponential inequality and the independence of the variables Z^k to obtain

$$\begin{aligned}
\mathbb{P}\left[W_p(\hat{\mu}_R^N, \hat{\mu}^N) > \varepsilon\right] &\leq \mathbb{P}\left[\frac{1}{N}\sum_{k=1}^N Z^k > \varepsilon^p\right] \\
&= \mathbb{P}\left[\exp\sum_{k=1}^N \theta(Z^k - \varepsilon^p) > 1\right] \\
&\leq \mathbb{E}\left(\exp\sum_{k=1}^N \theta(Z^k - \varepsilon^p)\right) \\
&= \exp\left(-N\left[\theta\varepsilon^p - \ln\mathbb{E}\exp(\theta Z_1)\right]\right).
\end{aligned} \tag{6.22}$$

In the case when <u>$p < 2$</u>, for any $\alpha_1 < \alpha < \dfrac{\lambda}{2}$, there exists some constant $R_0 = R_0(\alpha_1, p)$ such that

$$2^p\theta r^p \leq \alpha_1 r^2 + C,$$

for all $\theta > 0$ and $r \geq R_0\theta^{\frac{1}{2-p}}$, whence

$$\mathbb{E}\exp(\theta Z_1) \leq \mathbb{E}\exp(\alpha_1\,|X_1|^2\,\mathbf{1}_{|X^k|>R}) \leq 1 + E_\alpha e^{(\alpha_1-\alpha)R^2}.$$

As a consequence,

$$\mathbb{P}\left[W_p(\hat{\mu}_R^N, \hat{\mu}^N) > \varepsilon\right] \leq \exp\left(-N\left[\theta\,\varepsilon^p - E_\alpha e^{(\alpha_1-\alpha)R^2}\right]\right). \tag{6.23}$$

From (6.21), (6.23) and the triangular inequality for W_p,

$$\begin{aligned}
\mathbb{P}\left[W_p(\mu, \hat{\mu}^N) > \varepsilon\right] &\leq \mathbb{P}\left[W_p(\mu, \mu_R) + W_p(\mu_R, \hat{\mu}_R^N) + W_p(\hat{\mu}_R^N, \hat{\mu}^N) > \varepsilon\right] \\
&\leq \mathbb{P}\left[W_p(\mu_R, \hat{\mu}_R^N) > \eta\,\varepsilon - 2E_\alpha^{1/p} R\, e^{-\frac{\alpha}{p}R^2}\right] + \mathbb{P}\left[W_p(\hat{\mu}_R^N, \hat{\mu}^N) > (1-\eta)\varepsilon\right] \\
&\leq \mathbb{P}\left[W_p(\mu_R, \hat{\mu}_R^N) > \eta\,\varepsilon - 2E_\alpha^{1/p} R e^{-\frac{\alpha}{p}R^2}\right] \\
&\quad + \exp\left(-N\left(\theta(1-\eta)^p\varepsilon^p - E_\alpha\, e^{(\alpha_1-\alpha)\,R^2}\right)\right).
\end{aligned} \tag{6.24}$$

This estimate was established for any given $p \in [1,2)$, $\eta \in (0,1)$, $\varepsilon, \theta > 0$, $\alpha_1 < \alpha < \dfrac{\lambda}{2}$ and $R \geq \max\left(\sqrt{p/2\alpha}, R_0 \theta^{\frac{1}{2-p}}\right)$, where R_0 is a constant depending only on α_1 and p.

In the case when $\underline{p=2}$, we let $Z^k := |Y_k - X_k|^2\, \mathbf{1}_{|X_k|>R}$ $(k = 1, \ldots, N)$, and starting from inequality (6.22) again, we choose $\alpha_1 < \alpha$ and then $\theta := \alpha_1/2$: by definition of Z_1 and μ_R,

$$
\begin{aligned}
\mathbb{E}\left(\exp\left(\frac{\alpha_1}{2} Z_1\right)\right) &= \int_{\mathbb{R}^{2d}} \exp\left(\frac{\alpha_1}{2}|y-x|^2 \mathbf{1}_{|x|\geq R}\right) d\mu(x)\, d\mu_R(y) \\
&= \mu[B_R] + \frac{1}{\mu[B_R]} \int_{|y|\leq R} \int_{|x|\geq R} \exp\left(\frac{\alpha_1}{2}|y-x|^2\right) d\mu(x)\, d\mu(y) \\
&\leq 1 + (1 - E_\alpha e^{-\alpha R^2})^{-1} \int_{|y|\leq R} e^{\alpha_1 |y|^2}\, d\mu(y) \int_{|x|\geq R} e^{\alpha_1 |x|^2}\, d\mu(x) \\
&\leq 1 + 2\, E_\alpha^2\, e^{(\alpha_1 - \alpha)R^2}
\end{aligned}
$$

for R large enough, from which

$$
\mathbb{P}\left[W_2(\hat{\mu}_R^N, \hat{\mu}^N) > \varepsilon\right] \;\leq\; \exp\left(-N\left[\frac{\alpha_1}{2}\,\varepsilon^2 - 2\, E_\alpha^2\, e^{(\alpha_1-\alpha)R^2}\right]\right). \tag{6.25}
$$

To sum up, in the case $p = 2$ equation (6.24) writes

$$
\begin{aligned}
\mathbb{P}\left[W_2(\mu, \hat{\mu}^N) > \varepsilon)\right] \leq \mathbb{P}&\left[W_2(\mu_R, \hat{\mu}_R^N) > \eta\,\varepsilon - 2E_\alpha^{1/2} R e^{-\frac{\alpha}{2}R^2}\right] \\
&+ \exp\left(-N\left(\frac{\alpha_1}{2}(1-\eta)^2 \varepsilon^2 - 2\, E_\alpha^2\, e^{(\alpha_1-\alpha)\, R^2}\right)\right).
\end{aligned} \tag{6.26}
$$

So, apart from some error terms, for all $p \in [1,2]$ we have reduced the initial problem to establishing the result only for the probability law μ_R, whose support lies in the compact set B_R.

We end up this truncation procedure by proving that μ_R satisfies some modified T_p inequality. Let indeed ν be a probability measure on B_R, absolutely continuous with respect to μ (and hence with respect to μ_R) ; then, when R is larger than some constant depending only on E_α, we can write

$$
\begin{aligned}
H(\nu|\mu_R) - H(\nu|\mu) = \int_{B_R} \ln \frac{d\nu}{d\mu_R}\, d\nu - \int_{B_R} \ln \frac{d\nu}{d\mu}\, d\nu = \ln \mu[B_R] &\geq \ln\left(1 - E_\alpha e^{-\alpha R^2}\right) \\
&\geq -2E_\alpha e^{-\alpha R^2}.
\end{aligned} \tag{6.27}
$$

But μ satisfies a $T_p(\lambda)$ inequality, so

$$
H(\nu|\mu) \geq \frac{\lambda}{2} W_p^2(\mu, \nu) \geq \frac{\lambda}{2}\left(W_p(\mu_R, \nu) - W_p(\mu_R, \mu)\right)^2
$$

by triangular inequality. Combining this with (6.27), we obtain

$$
H(\nu|\mu_R) \geq \frac{\lambda}{2}\left(W_p(\mu_R, \nu) - W_p(\mu_R, \mu)\right)^2 - 2E_\alpha\, e^{-\alpha R^2}
$$

From this, inequality (6.21) and the elementary inequality

$$\forall a \in (0,1) \qquad \exists C_a > 0; \qquad \forall x, y \in \mathbb{R}, \qquad (x-y)^2 \geq (1-a)\, x^2 - C_a\, y^2, \tag{6.28}$$

we deduce that for any $\lambda_1 < \lambda$ there exists some constant K such that

$$H(\nu|\mu_R) \geq \frac{\lambda_1}{2} W_p(\mu_R, \nu)^2 - K R^2 e^{-\alpha R^2}. \tag{6.29}$$

Step 2 : Covering by small balls. In this second step we derive quantitative estimates on $\hat{\mu}_R^N$. Let ϕ be a bounded continuous function on B_R, and let $\mathcal{B}$ be a Borel set in $\mathcal{P}(B_R)$ (equipped with the weak topology of convergence against bounded continuous test functions). By Chebyshev's exponential inequality and the independence of the variables X_R^k,

$$\begin{aligned}
\mathbb{P}[\hat{\mu}_R^N \in \mathcal{B}] &\leq \exp\left(-N \inf_{\nu\in\mathcal{B}} \int_{B_R} \phi\, d\nu\right) \mathbb{E}\left(e^{N\int_{B_R} \phi\, d\hat{\mu}_R^N}\right) \\
&= \exp\left(-N \inf_{\nu\in\mathcal{B}} \left[\int_{B_R} \phi\, d\nu - \frac{1}{N}\log \mathbb{E}\left(e^{N\int_{B_R} \phi\, d\hat{\mu}_R^N}\right)\right]\right) \\
&= \exp\left(-N \inf_{\nu\in\mathcal{B}} \left[\int_{B_R} \phi\, d\nu - \frac{1}{N}\log \mathbb{E}\left(e^{\sum_{k=1}^N \phi(X_R^k)}\right)\right]\right) \\
&= \exp\left(-N \inf_{\nu\in\mathcal{B}} \left[\int_{B_R} \phi\, d\nu - \log \int_{B_R} e^{\phi} d\mu_R\right]\right).
\end{aligned}$$

As ϕ is arbitrary, we can pass to the supremum and find

$$\mathbb{P}[\hat{\mu}_R^N \in \mathcal{B}] \leq \exp\left(-N \sup_{\phi\in C_b(B_R)} \inf_{\nu\in\mathcal{B}} \left[\int_{B_R} \phi\, d\nu - \log \int_{B_R} e^{\phi}\, d\mu_R\right]\right).$$

Now we note that the quantity $\int \phi\, d\nu - \log \int e^\phi\, d\mu_R$ is concave in ϕ and linear continuous in ν ; if we further assume that $\mathcal{B}$ is convex and compact, then (for instance) Sion's min-max theorem [103, Theorem 4.2'] ensures that

$$\sup_{\phi\in C_b(B_R)} \inf_{\nu\in\mathcal{B}} \left[\int_{B_R} \phi\, d\nu - \log \int e^{\phi} d\mu_R\right] = \inf_{\nu\in\mathcal{B}} \sup_{\phi\in C_b(B_R)} \left[\int_{B_R} \phi\, d\nu - \log \int e^{\phi} d\mu_R\right].$$

By the dual formulation of the H functional [43, Lemma 6.2.13], we conclude that

$$\mathbb{P}[\hat{\mu}_R^N \in \mathcal{B}] \leq \exp\left(-N \inf_{\nu\in\mathcal{B}} H(\nu|\mu_R)\right). \tag{6.30}$$

Now, let $\delta > 0$ and let $\mathcal{A}$ be a measurable subset of $\mathcal{P}(B_R)$. We cover the latter with $\mathcal{N}^{\mathcal{A}}$ balls $(B_i)_{1\leq i\leq \mathcal{N}^{\mathcal{A}}}$ with radius $\delta/2$ in W_p metric. Each of these balls is convex and compact, and it is included in the δ-thickening of $\mathcal{A}$ in W_p metric, defined as

$$\mathcal{A}_\delta := \Big\{\nu \in \mathcal{P}(B_R); \quad \exists\, \nu_a \in \mathcal{A},\ W_p(\nu, \nu_a) \leq \delta\Big\}.$$

So, by (6.30) we get

$$\mathbb{P}[\hat{\mu}_R^N \in \mathcal{A}] \le \mathbb{P}\left[\hat{\mu}_R^N \in \bigcup_{i=1}^{\mathcal{N}^{\mathcal{A}}} B_i\right] \le \sum_{i=1}^{\mathcal{N}^{\mathcal{A}}} \mathbb{P}\left(\hat{\mu}_R^N \in B_i\right)$$
$$\le \sum_{i=1}^{\mathcal{N}^{\mathcal{A}}} \exp\left(-N \inf_{\nu \in B_i} H(\nu|\mu_R)\right) \le \mathcal{N}^{\mathcal{A}} \exp\left(-N \inf_{\nu \in \mathcal{A}_\delta} H(\nu|\mu_R)\right). \quad (6.31)$$

We now apply this estimate with

$$\mathcal{A} := \left\{\nu \in \mathcal{P}(B_R); \quad W_p(\nu, \mu_R) \ge \eta\varepsilon - 2E_\alpha^{1/p} R\, e^{-\frac{\alpha}{p}R^2}\right\}.$$

From (6.29) we have, for any $\nu \in \mathcal{A}_\delta$,

$$H(\nu|\mu_R) \ge \frac{\lambda_1}{2} W_p(\nu, \mu_R)^2 - KR^2 e^{-\alpha R^2} \ge \frac{\lambda_1}{2}\mu^2 - KR^2 e^{-\alpha R^2},$$

where

$$\mu := \max\left(\eta\, \varepsilon - 2E_\alpha^{1/p} R e^{-\frac{\alpha}{p}R^2} - \delta, 0\right).$$

Combining this with (6.31), we conclude that

$$\mathbb{P}\left[W_p(\mu_R, \hat{\mu}_R^N) \ge \eta\varepsilon - 2E_\alpha^{1/p} R e^{-\frac{\alpha}{p}R^2}\right] \le \mathcal{N}^{\mathcal{A}} \exp\left(-N\left[\frac{\lambda_1}{2}\mu^2 - KR^2 e^{-\alpha R^2}\right]\right). \quad (6.32)$$

Now, given any $\lambda_2 < \lambda_1$, it follows from (6.28) that there exist δ_1, η_1 and K_1, depending on $\alpha, \lambda_1, \lambda_2$, such that

$$\frac{\lambda_1}{2}\mu^2 - KR^2 e^{-\alpha R^2} \ge \frac{\lambda_2}{2}\varepsilon^2 - K_1 R^2 e^{-\alpha R^2} \quad (6.33)$$

where $\delta := \delta_1 \varepsilon$ and $\eta := \eta_1$.

Though this inequality holds independently of p, we shall use it only in the case when $p < 2$. In the case $p = 2$, on the other hand, we note that for any $\eta \in (0, 1)$,

$$\frac{\lambda_1}{2}\mu^2 - KR^2 e^{-\alpha R^2} \ge \frac{\lambda_2}{2}\eta^2\varepsilon^2 - K_1 R^2 e^{-\alpha R^2} \quad (6.34)$$

where $\delta := \delta_1 \varepsilon$.

Finally, we bound $\mathcal{N}^{\mathcal{A}}$ by means of Theorem 6.25 in Appendix 6.7 : there exists some constant C (only depending on d) such that for all $R > 0$ and $\delta > 0$ the set $\mathcal{P}(B_R)$ can be covered by

$$\left(C\frac{R}{\delta} \vee 1\right)^{\left(C\frac{R}{\delta}\right)^d}$$

balls of radius δ in W_p metric, where $a \vee b$ stands for $\max(a,b)$. In particular, given $\delta = \delta_1 \varepsilon$, we can choose

$$\mathcal{N}^{\mathcal{A}} \leq \left(K_2 \frac{R}{\varepsilon} \vee 1 \right)^{\left(K_2 \frac{R}{\varepsilon}\right)^d} \tag{6.35}$$

balls of radius δ, for some constant K_2 depending on λ_1 and λ_2 (via δ_1) but neither on ε nor on R. (The purpose of the 1 in $(K_2 R/\varepsilon \vee 1)$ is to make sure that the estimate is also valid when $\varepsilon > R$.)

Combining (6.24), (6.32), (6.33) and (6.35), we find that, given $p \in [1,2)$, $\lambda_2 < \lambda$ and $\alpha_1 < \alpha < \dfrac{\lambda}{2}$, there exist some constants K_1, K_2, K_3 and R_1 such that for all $\varepsilon, \zeta > 0$ and $R \geq R_1 \max(1, \zeta^{\frac{1}{2-p}})$,

$$\mathbb{P}\left[W_p(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \left(K_2 \frac{R}{\varepsilon} \vee 1 \right)^{K_2 \left(\frac{R}{\varepsilon}\right)^d} \exp\left(-N \left[\frac{\lambda_2 \varepsilon^2}{2} - K_1 R^2 e^{-\alpha R^2} \right] \right) \\ + \exp\left(-N \left(K_3\, \zeta\, \varepsilon^p - K_4 e^{(\alpha_1 - \alpha)\, R^2} \right) \right) \tag{6.36}$$

for some constant $K_4 = K_4(\theta, \alpha_1)$. In the case when $\underline{p = 2}$, we obtain similarly

$$\mathbb{P}\left[W_2(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \left(K_2 \frac{R}{\varepsilon} \vee 1 \right)^{K_2 \left(\frac{R}{\varepsilon}\right)^d} \exp\left(-N \left[\frac{\lambda_2}{2} \eta^2 \varepsilon^2 - K_1 R^2 e^{-\alpha R^2} \right] \right) \\ + \exp\left(-N \left(\frac{\alpha_1}{2} (1-\eta)^2\, \varepsilon^2 - K_4 e^{(\alpha_1 - \alpha)\, R^2} \right) \right) \tag{6.37}$$

for any $\eta \in (0,1)$ and $R \geq R_1$.

These estimates are not really appealing (!), but they are rather precise and general. In the rest of the section we shall show that an adequate choice of R leads to a simplified expression.

Step 3 : Choice of the parameters.

We first consider the case when $\underline{p \in [1,2)}$. Let $\lambda' < \lambda_2$, $\alpha' < \alpha$ and $d_1 > d$. We claim that

$$\mathbb{P}\left[W_p(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \exp\left(-\frac{\lambda'}{2} N\, \varepsilon^2 \right) + \exp\left(-\alpha' N\, \varepsilon^2 \right)$$

as soon as

$$R^2 \geq R_2 \max\left(1, \varepsilon^2, \ln\left(\frac{1}{\varepsilon^2}\right)\right), \qquad N\, \varepsilon^{d_1+2} \geq K_5 R^{d_1} \tag{6.38}$$

for some constants R_2 and K_5 depending on μ only through λ, α and E_α.

Indeed, on one hand

$$K_2 \left(\frac{R}{\varepsilon}\right)^d \ln\left(K_2 \frac{R}{\varepsilon}\right) \leq K_6 \left(\frac{R}{\varepsilon}\right)^{d_1}$$

for some constant K_6, on the other hand

$$K_1 R^2 e^{-\alpha R^2} \le e^{-\alpha_1 R^2}$$

for R large enough, and then

$$K_6 \left(\frac{R}{\varepsilon}\right)^{d_1} - N\left[\frac{\lambda_2 \varepsilon^2}{2} - e^{-\alpha_1 R^2}\right] \le -N\,\frac{\lambda' \varepsilon^2}{2}$$

for $R^2/\ln(\frac{1}{\varepsilon^2})$ and $N\,\varepsilon^{d_1+2}/R^{d_1}$ large enough ; this is enough to bound the first term in the right-hand side of (6.36) if moreover R/ε is large enough.

Moreover, letting $\alpha_2 \in (\alpha', \alpha_1)$, we can choose ζ in such a way that $K_3\zeta = \varepsilon^{2-p}$, so that

$$\exp\left(-N\left(K_3\zeta\varepsilon^p - K_4\, e^{(\alpha_1-\alpha)\,R^2}\right)\right) = \exp\left(-N\left(\alpha_2\,\varepsilon^2 - K_4 e^{(\alpha_1-\alpha)\,R^2}\right)\right),$$

which in the end can be bounded by

$$\exp\left(-N\,\alpha'\varepsilon^2\right)$$

if R and $R^2/\ln(\frac{1}{\varepsilon^2})$ are large enough. With this one can get a bound on the right-hand side of (6.36).

Now let us check that conditions (6.38) can indeed be fulfilled. Clearly, the first condition holds true for all $\varepsilon \in (0,1)$ and $R^2 \ge R_3 \ln(\frac{K_6}{\varepsilon^2})$, where R_3 and K_6 are positive constants. Then, we can choose

$$R := \left(\frac{N}{K_5}\varepsilon^{d_1+2}\right)^{1/d_1}$$

so that the second condition holds as an equality. This choice is admissible as soon as

$$\left(\frac{N}{K_5}\,\varepsilon^{d_1+2}\right)^{2/d_1} \ge R_3 \ln\left(\frac{K_5}{\varepsilon^2}\right)$$

and this, in turn, holds true as soon as

$$N \ge K_7\;\varepsilon^{-(d'+2)}, \tag{6.39}$$

where d' is such that $d' > d$, and K_7 is large enough.

If $\varepsilon \ge 1$, then we can choose $R^2 = R_2\varepsilon^2$, i.e. $R = \sqrt{R_2}\varepsilon$, and then the second inequality in (6.38) will be true as soon as N is large enough.

To sum up : Given $d' > d$, $\lambda' < \lambda$ and $\alpha' < \alpha$, there exists some constant N_0, depending on d' and depending on μ only through λ, α and E_α, such that for all $\varepsilon > 0$,

$$\mathbb{P}\left[W_p(\mu, \hat\mu^N) > \varepsilon\right] \le \exp\left(-\frac{\lambda'}{2}\,N\,\varepsilon^2\right) + \exp\left(-\alpha'\,N\,\varepsilon^2\right)$$

as soon as $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$. Then we note that, given $K < \min\left(\frac{\lambda'}{2}, \alpha'\right)$, the inequality

$$\exp\left(-\frac{\lambda'}{2} N \varepsilon^2\right) + \exp\left(-\alpha' N \varepsilon^2\right) \leq \exp\left(-K N \varepsilon^2\right)$$

holds if condition (6.39) is satisfied for some K_7 large enough. To conclude the proof of Theorem 6.1 in the case when $p \in [1, 2)$, it is sufficient to choose $\lambda' < \lambda$, $\alpha < \lambda/2$.

Now, in the case when $\underline{p = 2}$, given $\lambda_3 < \lambda_2$ and $\alpha_2 < \alpha_1$, conditions (6.38) imply

$$\mathbb{P}\left[W_2(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \exp\left(-\frac{\lambda_3}{2}\eta^2 N\varepsilon^2\right) + \exp\left(-\frac{\alpha_2}{2}(1-\eta)^2 N\varepsilon^2\right).$$

Then we let $\alpha_2 := \frac{\lambda_3}{2}$ and $\eta := \sqrt{2} - 1$, so that

$$\frac{\lambda_3}{2}\eta^2 = \frac{\alpha_2}{2}(1-\eta)^2.$$

Then

$$\mathbb{P}\left[W_2(\mu, \hat{\mu}^N) > \varepsilon\right] \leq 2 \exp\left(-(3 - 2\sqrt{2})\frac{\lambda_3}{2} N \varepsilon^2\right);$$

for $\lambda' < \lambda$, the above quantity is bounded by

$$\exp\left(-(3 - 2\sqrt{2})\frac{\lambda'}{2} N \varepsilon^2\right)$$

as soon as (6.39) is enforced with K_7 large enough. This concludes the argument. $\square$

6.2.2 Proof of Theorem 6.5

It is very similar to the proof of Theorem 6.1, so we shall only explain where the differences lie. Obviously, the main difficulty will consist in the control of tails.

We first let $p \in [1, q)$, $\alpha \in [1, \frac{q}{p})$ and $R > 0$, and introduce

$$M_q := \int_{\mathbb{R}^d} |x|^q \, d\mu(x).$$

Then (6.21) may be replaced by

$$W_p^p(\mu, \mu_R) \leq 2^p M_q R^{p-q}, \tag{6.40}$$

and (6.22) by

$$\mathbb{P}\left[W_p(\hat{\mu}_R^N, \hat{\mu}^N) > \varepsilon\right] \leq C\, N^{\overline{\alpha}-\alpha} \frac{R^{\alpha p - q}}{(\varepsilon^p - C\, R^{p-q})^\alpha} \tag{6.41}$$

for some constant C depending on α and M_q.

Let us establish for instance (6.41). Introduce

$$Z_k = |Y_k - X_k|^p \, \mathbf{1}_{|X_k|>R} \qquad (1 \le k \le N).$$

By Chebychev's inequality,

$$\mathbb{P}\left[W_p(\hat\mu_R^N, \hat\mu^N) > \varepsilon\right] \le \mathbb{P}\left[\frac{1}{N}\sum_{k=1}^N Z_k > \varepsilon^p\right] = \mathbb{P}\left[\frac{1}{N}\sum_{k=1}^N (Z_k - \mathbb{E}\,Z_k) > \varepsilon^p - \mathbb{E}\,Z_1\right]$$
$$\le \frac{\mathbb{E}\left|\sum_{k=1}^N (Z_k - \mathbb{E}\,Z_k)\right|^\alpha}{(N\,(\varepsilon^p - \mathbb{E}\,Z_1))^\alpha}$$

provided that $\varepsilon^p > \mathbb{E}\,Z_1$. But, since the random variables $(Z_k - \mathbb{E}\,Z_k)_k$ are independent and identically distributed, with zero mean, there exists some constant C depending on α such that

$$\mathbb{E}\left|\sum_{k=1}^N (Z_k - \mathbb{E}\,Z_k)\right|^\alpha \le C\,N^{\overline{\alpha}}\,\mathbb{E}|Z_1 - \mathbb{E}\,Z_1|^\alpha$$

where $\overline{\alpha} := \max(\alpha/2, 1)$. This inequality is a consequence of Rosenthal's inequality in the case when $\alpha \ge 2$, but also holds true if $\alpha \in [1,2)$ (see for instance [90, pp. 62 and 82]). Then, on one hand,

$$\mathbb{E}\,Z_1 = \mathbb{E}\,|Y_1 - X_1|^p\,\mathbf{1}_{|X_1|>R} \le 2^p M_q R^{p-q},$$

while on the other hand,

$$\mathbb{E}|Z_1 - \mathbb{E}Z_1|^\alpha = \mathbb{E}\left||Y_1 - X_1|^p\,\mathbf{1}_{|X_1|>R} - \mathbb{E}|Y_1 - X_1|^p\,\mathbf{1}_{|X_1|>R}\right|^\alpha$$
$$\le C\mathbb{E}|Y_1 - X_1|^{\alpha p}\,\mathbf{1}_{|X_1|>R} \le C\,M_q\,R^{\alpha p - q}$$

with C standing for various constants. Collecting these two estimates, we conclude to the validity of (6.41) for $R^{q-p}\,\varepsilon^p$ large enough.

Then (6.40) and (6.41) together ensure that

$$\mathbb{P}[W_p(\mu, \hat\mu^N) > \varepsilon] \le \mathbb{P}\left[W_p(\mu_R, \hat\mu_R^N) > \eta\,\varepsilon - 2M_q^{1/p}R^{1-q/p}\right]$$
$$+ \, C\,N^{\overline{\alpha}-\alpha}\frac{R^{\alpha\,p-q}}{((1-\eta)^p\varepsilon^p - C\,R^{p-q})^\alpha} \qquad (6.42)$$

for any $\varepsilon \in (0,1)$, $\eta > 0$ and $R^{q-p}\,\varepsilon^p\,(1-\eta)^p$ large enough.

Since μ_R is supported in B_R, the Csiszár-Kullback-Pinsker inequality and Kantorovich-Rubinstein formulation of the W_1 distance together ensure that it satisfies a $T_1(R^{-2})$ inequality (see e.g. [23, Particular Case 5] with $p = 1$). This estimate also extends to any W_p distance, not as a penalized T_p inequality as in (6.29), but rather as

$$W_p^{2p}(\nu, \mu_R) \le 2^{2p-1}R^{2p}H(\nu|\mu_R) \qquad (6.43)$$

(see again [23, Particular Case 5]).

From (6.42) and (6.43) we deduce (as in (6.37)) that

$$\mathbb{P}\left[W_p(\hat{\mu}^N,\mu) > \varepsilon\right] \le \left(K_1 \frac{R}{\delta}\right)^{K_1\left(\frac{R}{\delta}\right)^d} \exp\left(-\frac{N\mu^{2p}}{2^{2p-1}R^{2p}}\right) + C\,N^{\overline{\alpha}-\alpha}\frac{R^{\alpha\,p-q}}{((1-\eta)^p\varepsilon^p - C\,R^{p-q})^{\alpha}} \tag{6.44}$$

for any δ, where now

$$\mu := \left(\eta\varepsilon - 2M^{1/p}R^{1-q/p} - \delta\right)^+.$$

Letting $\eta_1 < \eta$ and $d' > d$, and choosing $\delta = \delta_0\,\varepsilon$, we deduce

$$\mathbb{P}\left[(W_p(\hat{\mu}^N,\mu) > \varepsilon\right] \le \exp\left(\left(\frac{R}{\varepsilon}\right)^{d'} - \frac{\eta_1^{2p}}{2^{2p-1}}\frac{N\,\varepsilon^{2p}}{R^{2p}} + \frac{K_1}{2^{2p-1}}\frac{N}{R^{2q}}\right) + C\,N^{\bar{\alpha}-\alpha}\frac{R^{\alpha\,p-q}}{((1-\eta_1)^p\varepsilon^p - C\,R^{p-q})^{\alpha}}$$

for $R^{q-p}\,\varepsilon^p\,(1-\eta_1)^p$ large enough, and then

$$\mathbb{P}\left[W_p(\hat{\mu}^N,\mu) > \varepsilon\right] \le \exp\left(-\frac{\eta_2^{2p}}{2^{2p-1}}\frac{N\,\varepsilon^{2p}}{R^{2p}}\right) + C\,N^{\overline{\alpha}-\alpha}\frac{R^{\alpha\,p-q}}{(1-\eta_2)^{\alpha p}\varepsilon^{\alpha p}} \tag{6.45}$$

for $\eta_2 < \eta_1$, provided that the conditions

$$R \ge R_1\varepsilon^{-\frac{p}{q-p}}, \qquad N \ge K_2\left(\frac{R}{\varepsilon}\right)^{2p+d'} \tag{6.46}$$

hold for some R_1 and K_2.

Given any choice of R as a product of powers of N and ε, the first term in the right-hand side of (6.45) will always be smaller than the second one, if N goes to infinity while ε is kept fixed ; thus we can choose R minimizing the second term under the above conditions. Then the second condition in (6.46) will be fulfilled as an equality :

$$R = K_3\,\varepsilon\,N^{\frac{1}{2p+d'}}.$$

As for the first condition in (6.46), it can be rewritten as

$$N \ge N_0\,\varepsilon^{-q\frac{2p+d'}{q-p}},$$

and then, by (6.45),

$$\mathbb{P}\left[W_p(\hat{\mu}^N,\mu) > \varepsilon\right] \le \exp\left(-K_5 N^{\frac{d'}{2p+d'}}\right) + K_6\,\varepsilon^{-q}\,N^{\overline{\alpha}-\alpha+\frac{\alpha p-q}{2p+d'}}.$$

Hence

$$\mathbb{P}\left[W_p(\hat{\mu}^N,\mu) > \varepsilon\right] \le \varepsilon^{-q}\,N^{\overline{\alpha}-\alpha} \tag{6.47}$$

for all $\varepsilon \in (0,1)$ and N larger than some constant and, given $d' > d$, for all $\varepsilon \geq 1$ and $N \geq M\varepsilon^{d'-d}$ where M is large enough.

In the first case when $p \geq q/2$, any admissible α belongs to $[1, q/p) \subset [1,2]$, so $\overline{\alpha} = 1$. If $\delta \in (0, q/p-1)$, we get from (6.47), with $\alpha = q/p - \delta$, that

$$\mathbb{P}\Big[W_p(\hat{\mu}^N, \mu) > \varepsilon\Big] \leq \varepsilon^{-q} N^{1-q/p+\delta}$$

for all $\varepsilon > 0$ and

$$N \geq N_0 \max\big(\varepsilon^{-q\frac{2p+d'}{q-p}}, \varepsilon^{d'-d}\big).$$

In the second case when $p < q/2$, we only consider admissible α's in $[2, q/p) \subset [1, q/p)$, so that $\overline{\alpha} - \alpha = -\alpha/2$. Choosing $\delta \in (0, q/p-2)$, we get from (6.47)

$$\mathbb{P}\Big[W_p(\hat{\mu}^N, \mu) > \varepsilon\Big] \leq \varepsilon^{-q} N^{-q/2p+\delta/2}$$

under the same conditions on N as before. This concludes the argument. □

6.2.3 Proof of Theorem 6.6

It is again based on the same principles as the proofs of Theorems 6.1 and 6.5, with the help of functional inequalities investigated in [23] and [39]. We skip the argument, which the reader can easily reconstruct by following the same lines as above. □

6.2.4 Data reconstruction estimates

Finally, we show how the above concentration estimates imply data reconstruction estimates. This is a rather general estimate, which is treated here along the lines of [98, Section 5] and [111, Problem 10].

Proposition 6.12. *Let μ be a probability measure on $\mathbb{R}^d$, with density f with respect to Lebesgue measure. Let $X_1, \ldots, X_N$ be random points in $\mathbb{R}^d$, and let ζ be a Lipschitz, nonnegative kernel with unit integral. Define the random measure $\hat{\mu}$ and the random function $\hat{f}_{\zeta,\alpha}$ by*

$$\hat{\mu} := \frac{1}{N}\sum_{i=1}^N \delta_{X_i}, \qquad \hat{f}_{\zeta,\alpha}(x) := \frac{1}{N}\sum_{i=1}^N \zeta_\alpha(x - X_i), \qquad \zeta_\alpha(x) = \frac{1}{\alpha^d}\zeta\left(\frac{x}{\alpha}\right).$$

Then,

$$\sup_{x\in\mathbb{R}^d} |\hat{f}_{\zeta,\alpha}(x) - f(x)| \leq \frac{\|\zeta\|_{\mathrm{Lip}}}{\alpha^{d+1}} W_1(\hat{\mu}, \mu) + \delta(\alpha), \tag{6.48}$$

where δ stands for the modulus of continuity of f, defined as

$$\delta(\varepsilon) := \sup_{|x-y|\leq\varepsilon} |f(x) - f(y)|.$$

As a consequence, if f is Lipschitz, then there exist some constants $a, K > 0$, only depending on d, $\|f\|_{\mathrm{Lip}}$ and $\|\zeta\|_{\mathrm{Lip}}$, such that

$$\mathbb{P}\Big[\|\hat{f}_{\zeta,a\varepsilon} - f\|_{L^\infty} > \varepsilon\Big] \le \mathbb{P}\Big[W_1(\hat{\mu},\mu) > K\varepsilon^{d+2}\Big] \tag{6.49}$$

for all $\varepsilon > 0$.

Proof. First,

$$|\mu * \zeta_\alpha(x) - f(x)| = \left|\int_{\mathbb{R}^d} \zeta_\alpha(x-y)\,\big(f(y)-f(x)\big)\,dy\right| \le \int_{\mathbb{R}^d} \zeta_\alpha(x-y)|f(y)-f(x)|\,dy.$$

Since $\zeta_\alpha(x-y)$ is supported in $\{|x-y| \le \alpha\}$, and ζ_α is a probability density, we deduce

$$|\mu * \zeta_\alpha(x) - f(x)| \le \delta(\alpha). \tag{6.50}$$

Now, if x is some point in $\mathbb{R}^d$, then, thanks to the Kantorovich-Rubinstein dual formulation (6.8),

$$\Big|\hat{f}_{\zeta,\alpha} - \mu * \zeta_\alpha\Big|(x) = \left|\int_{\mathbb{R}^d} \zeta_\alpha(x-y)\,d[\hat{\mu}-\mu](y)\right| \le \|\zeta_\alpha(x-\cdot)\|_{\mathrm{Lip}} W_1(\hat{\mu},\mu) = \frac{\|\zeta\|_{\mathrm{Lip}}}{\alpha^{d+1}} W_1(\hat{\mu},\mu).$$

To conclude the proof of (6.48), it suffices to combine this bound with (6.50).

Now, let $L := \max(\|f\|_{\mathrm{Lip}}, \|\zeta\|_{\mathrm{Lip}})$, and $\alpha := \varepsilon/(2L)$. The bound (6.48) turns into

$$\|\hat{f}_{\zeta,\alpha} - f\|_{L^\infty} \le L\big(\frac{W_1(\hat{\mu},\mu)}{\alpha^{d+1}} + \alpha\big) \le \left(\frac{(2L)^{d+1}L}{\varepsilon^{d+1}}\right) W_1(\hat{\mu},\mu) + \frac{\varepsilon}{2}.$$

In particular,

$$\mathbb{P}\Big[\|\hat{f}_{\zeta,\alpha} - f\|_{L^\infty} > \varepsilon\Big] \le \mathbb{P}\left[W_1(\hat{\mu},\mu) > \frac{\varepsilon^{d+2}}{(2L)^{d+2}}\right],$$

which is estimate (6.49). □

Remark 6.13. Estimate (6.49), combined with Theorem 6.1 or Theorem 6.5, yields simple quantitative (non-asymptotic) deviation inequalities for empirical distribution functions in supremum norm. We refer to Gao [53] for a recent study of deviation inequalities for empirical distribution functions, both in moderate and large deviations regimes.

6.3 PDE estimates

Now we start the study of our model system for interacting particles. The first step towards our proof of Theorem 6.7 consists in deriving suitable a priori estimates on the solution to the nonlinear limit partial differential equation (6.16). In this section, we recall some estimates which have already been established by various authors, and derive some new ones. All estimates will be effective.

6.3.1 Notation

In the sequel, μ_0 is a probability measure, taken as an initial datum for equation (6.16), and various regularity assumptions will later be made on μ_0. Assumptions (6.17) will always be made on V and W, even if they are not recalled explicitly ; we shall only mention additional regularity assumptions, when used in our estimates. Moreover, we shall write

$$\Gamma := \max(|\gamma|, |\gamma'|). \tag{6.51}$$

The notation μ_t will always stand for the solution (unique under our assumptions) of (6.16).

We also write

$$e(t) := \int_{\mathbb{R}^d} |x|^2 \, d\mu_t(x)$$

for the (kinetic) energy associated with μ_t, and

$$M_\alpha(t) := \int_{\mathbb{R}^d} e^{\alpha|x|^2} \, d\mu_t(x)$$

for the square exponential moment of order α.

The scalar product between two vectors $v, w \in \mathbb{R}^d$ will be denoted by $v \cdot w$. The symbols C and K will often be used to denote various positive constants ; in general what will matter is an upper bound on constants denoted C, and a lower bound on constants denoted K. The space $\mathcal{C}^k$ is the space of k times differentiable continuous functions.

6.3.2 Decay at infinity

In this subsection, we prove the propagation of strong decay estimates at infinity :

Proposition 6.14. *With the conventions of Subsection 6.3.1, let $\overline{\eta}$ be $-\gamma$ if $\gamma < 0$, and an arbitrary negative number otherwise. Let*

$$a := 2(\beta + \overline{\eta}), \qquad \overline{G} := 2\,d + \frac{|\nabla V(0)|^2}{2\,|\overline{\eta}|}.$$

Then

(i) $e(t) \le e^{-at} \left[e(0) + \overline{G}\, \frac{e^{at} - 1}{a} \right]$ *;*

(ii) For any $\alpha_0 > 0$ there is a continuous positive function $\alpha(t)$ such that $\alpha(0) = \alpha_0$ and

$$M_{\alpha_0}(0) < +\infty \Longrightarrow M_{\alpha(t)}(t) < +\infty. \tag{6.52}$$

(iii) Moreover, in the "uniformly convex case" when $\beta > 0$ and $\beta + \gamma > 0$, then there is $\alpha > 0$ such that

$$\sup_{t \ge 0} e(t) < +\infty, \qquad \sup_{t \ge 0} M_\alpha(t) < +\infty.$$

Corollary 6.15. *If μ_0 admits a finite square exponential moment, then μ_t satisfies $T_1(\lambda_t)$, for some function $\lambda_t > 0$, bounded below on any interval $[0,T]$ $(T < \infty)$.*

Proof. We start with (i). For simplicity we shall pretend that μ_t is a smoothly differentiable function of t, with rapid decay, so that all computations based on integrating equation (6.16) against $|x|^2$ are justified. These assumptions are not a priori satisfied, but the resulting bounds can easily be rigorously justified with standard but tedious approximation arguments. With that in mind, we compute

$$e'(t) = 2\,d - 2\int_{\mathbb{R}^d} (x \cdot \nabla V(x) + x \cdot \nabla W * \mu_t(x))\, d\mu_t(x)$$

with

$$-2\int_{\mathbb{R}^d} x \cdot \nabla V(x)\, d\mu_t(x) \le -2\beta \int_{\mathbb{R}^d} |x|^2\, d\mu_t(x) - 2\nabla V(0) \cdot \int_{\mathbb{R}^d} x\, d\mu_t(x).$$

Since ∇W is an odd function, we have

$$\begin{aligned}
-2\int_{\mathbb{R}^d} x \cdot \nabla W * \mu_t(x)\, d\mu_t(x) &= -2\iint x \cdot \nabla W(x-y)\, d\mu_t(y)\, d\mu_t(x)\\
&= -\iint (x-y) \cdot \nabla W(x-y)\, d\mu_t(y)\, d\mu_t(x)\\
&\le -\gamma \iint |x-y|^2\, d\mu_t(y)\, d\mu_t(x)\\
&= -2\,\gamma \left[\int |x|^2\, d\mu_t(x) - \left|\int x\, d\mu_t(x)\right|^2\right].
\end{aligned}$$

If $\gamma < 0$, then

$$\begin{aligned}
e'(t) &\le 2\,d - 2(\gamma+\beta)e(t) + 2\,\gamma\left|\int x\, d\mu_t(x) + \frac{\nabla V(0)}{2\,|\gamma|}\right|^2 + \frac{|\nabla V(0)|^2}{2\,|\gamma|}\\
&\le 2\,d - 2(\gamma+\beta)e(t) + \frac{|\nabla V(0)|^2}{2\,|\gamma|},
\end{aligned}$$

and if $\gamma \ge 0$, then for any $\overline{\eta} < 0$

$$\begin{aligned}
e'(t) &\le 2\,d - 2(\overline{\eta}+\beta)e(t) - 2\gamma\left(\int |x|^2\, d\mu_t(x) - \left|\int x\, d\mu_t(x)\right|^2\right) + \frac{|\nabla V(0)|^2}{2\,|\overline{\eta}|}\\
&\le 2\,d - 2(\overline{\eta}+\beta)e(t) + \frac{|\nabla V(0)|^2}{2\,|\overline{\eta}|}.
\end{aligned}$$

This leads to

$$e'(t) \le \overline{G} - a\,e(t),$$

and the conclusion follows easily by Gronwall's lemma.

We now turn to (ii). Let α be some arbitrary nonnegative $\mathcal{C}^1$ function on $\mathbb{R}_+$. By using the equation (6.16), we compute

$$\frac{d}{dt}\int e^{\alpha(t)|x|^2}\,d\mu_t(x)=\int\big[2d\alpha+4\alpha^2|x|^2-2\alpha x\cdot\nabla V(x)-2\alpha x\cdot\nabla W*\mu_t(x)+\alpha'(t)|x|^2\big]e^{\alpha|x|^2}\,d\mu_t(x).$$

Since $D^2V(x)\geq\beta I$ for all $x\in\mathbb{R}^d$, we can write

$$-x\,\cdot\,\nabla V(x)\leq -x\,\cdot\,\nabla V(0)-\beta|x|^2\leq-\beta|x|^2+|\nabla V(0)||x|\leq(\delta-\beta)|x|^2+\frac{C}{4\,\delta} \tag{6.53}$$

for any $\delta>0$ and $x\in\mathbb{R}^d$.

Next, our assumptions on W imply $\nabla W(0)=0$, and $\gamma I\leq D^2W(x)\leq\gamma' I$, so

$$x\,\cdot\,\nabla W(x)\geq\gamma\,|x|^2\qquad\text{and}\qquad|x\,\cdot\,D^2W(z)\,y|\leq\Gamma|x|\,|y|$$

for all $x,y,z\in\mathbb{R}^d$, with Γ defined by (6.51). Hence, by Taylor's formula,

$$\begin{aligned}
-x\,\cdot\,\nabla W*\mu_t(x) &= -\int_{\mathbb{R}^d}x\,\cdot\,\nabla W(x-y)\,d\mu_t(y)\\
&= -x\,\cdot\,\nabla W(x)+\int_{\mathbb{R}^d}\int_0^1 x\,\cdot\,D^2W(x-s\,y)\,y\,d\mu_t(y)\,ds\\
&\leq -\gamma\,|x|^2+\Gamma\,|x|\int_{\mathbb{R}^d}|y|\,d\mu_t(y)\\
&\leq (-\gamma+\Gamma\eta)|x|^2+\frac{\Gamma}{4\,\eta}\,e(t),
\end{aligned}\tag{6.54}$$

where η is any positive number.

From (6.53) and (6.54) we obtain

$$\frac{d}{dt}\Big(M_{\alpha(t)}(t)\Big)\leq\int_{\mathbb{R}^d}[A(t)+B(t)|x|^2]\,e^{\alpha(t)|x|^2}\,d\mu_t(x) \tag{6.55}$$

where

$$A(t)=C\alpha(t)\,(1+e(t)),\qquad B(t)=\alpha'(t)+4\,\alpha(t)^2+b\,\alpha(t),$$

and C is a finite constant, while $b=-2(\gamma+\beta-\delta-\Gamma\eta)$.

We now choose $\alpha(t)$ in such a way that $B(t)\equiv0$, i.e.

$$\alpha'(t)+4\,\alpha^2(t)+b\,\alpha(t)=0,\qquad\alpha(0)=\alpha_0.$$

This integrates to

$$\alpha(t)=e^{-b\,t}\left(\frac{1}{\alpha_0}+4\,\frac{1-e^{-b\,t}}{b}\right)^{-1}\qquad\left(=\left(\frac{1}{\alpha_0}+4\,t\right)^{-1}\ \text{ if }\ b=0\right).$$

Obviously α is a continuous positive function, and our estimates imply

$$\frac{d}{dt}\Big(M_{\alpha(t)}(t)\Big) \leq A(t) M_{\alpha(t)}(t).$$

We conclude by using Gronwall's lemma that

$$M_{\alpha(t)}(t) \leq \exp\left(\int_0^t A(s)\,ds\right) M_{\alpha_0}(0).$$

Next, the estimate (iii) for $e(t)$ is an easy consequence of our explicit estimates when $\beta > 0, \beta+\gamma > 0$ (in the case when $\gamma \geq 0$ and $\beta > 0$, we choose $\overline{\eta} \in (0,\beta)$).

As for the estimate about $M_\alpha(t)$, it will result from a slightly more precise computation. From (6.55), we have

$$\frac{d}{dt}\int_{\mathbb{R}^d} e^{\alpha\,|x|^2}\,d\mu_t(x) \leq \int_{\mathbb{R}^d} [A(t) + B|x|^2]\,e^{\alpha|x|^2}\,d\mu_t(x) \tag{6.56}$$

where A is bounded on $\mathbb{R}_+$ by some constant a, and

$$B = 2\,\alpha\,[2\alpha - (\beta+\gamma-\delta-\Gamma\eta)].$$

Since $\beta+\gamma > 0$, for any fixed α in $\left(0\,,\dfrac{\beta+\gamma}{2}\right)$ we can choose $\delta, \eta > 0$ such that $B < 0$. Letting $R^2 = -a/B$ and $G = -B > 0$, equation (6.56) becomes

$$\frac{d}{dt}\int_{\mathbb{R}^d} e^{\alpha\,|x|^2}\,d\mu_t(x) \leq G\int_{\mathbb{R}^d} (R^2 - |x|^2)\,e^{\alpha\,|x|^2}\,d\mu_t(x). \tag{6.57}$$

Let $p > 1$. The formula

$$\begin{aligned}\int_{|x|>p\,R} (R^2-|x|^2)\,e^{\alpha\,|x|^2}\,d\mu_t(x) &\leq R^2(1-p^2)\int_{|x|>p\,R} e^{\alpha\,|x|^2}\,d\mu_t(x)\\ &= R^2(1-p^2)\left[\int_{\mathbb{R}^d} e^{\alpha\,|x|^2}\,d\mu_t(x) - \int_{|x|\leq p\,R} e^{\alpha\,|x|^2}\,d\mu_t(x)\right]\end{aligned}$$

leads to

$$\int_{\mathbb{R}^d} (R^2-|x|^2)\,e^{\alpha\,|x|^2}\,d\mu_t(x) \leq \int_{|x|\leq p\,R} (R^2p^2-|x|^2)\,e^{\alpha\,|x|^2}\,d\mu_t(x) + R^2(1-p^2)M_\alpha$$

by decomposing the integral on the sets $\{|x| \leq p\,R\}$ and $\{|x| > p\,R\}$. From (6.57) we deduce

$$(M_\alpha)'(t) + \omega_1\,M_\alpha(t) \leq \omega_2$$

where ω_1 and ω_2 are positive constants. It follows that $M_\alpha(t)$ remains bounded on $\mathbb{R}_+$ if $M_\alpha(0) < +\infty$, and this concludes the argument. □

6.3.3 Time-regularity

Now we study the time-regularity of μ_t.

Proposition 6.16. *With the conventions of Subsection 6.3.1, for any $T < +\infty$ there exists a constant $C(T)$ such that*

$$\forall s,t \in [0,T], \qquad W_1(\mu_t, \mu_s) \;\le\; C(T)\, |t-s|^{1/2}. \tag{6.58}$$

Remark 6.17. The exponent $1/2$ is natural in small time if no regularity assumption is made on μ_0 ; it can be improved if t, s are assumed to be bounded below by some $t_0 > 0$. Also, in view of the results of convergence to equilibrium recalled later on, the constant $C(T)$ might be chosen independent of T if $\beta > 0, \beta + 2\,\gamma > 0$.

Remark 6.18. A stochastic proof of (6.58) is possible, via the study of continuity estimates for Y_t, which in any case will be useful later on. But here we prefer to present an analytical proof, to stress the fact that estimates in this section are purely analytical statements.

Proof. Let L be the linear operator $-\Delta - \nabla\cdot(\cdot\nabla V + \nabla(W*\mu_t))$, and let e^{-tL} be the associated semigroup : from our assumptions and estimates it follows that it is well-defined, at least for initial data which admit a finite square exponential moment. Of course $\mu_t = e^{-tL}\mu_0$. It follows that

$$\begin{aligned} W_2(\mu_s, \mu_t) = W_2(\mu_s, e^{-(t-s)L}\mu_s) &= W_2\left(\int_{\mathbb{R}^d} \delta_y \, d\mu_s(y),\; \int_{\mathbb{R}^d} e^{-(t-s)L}\delta_y \, d\mu_s(y)\right) \\ &\le \int_{\mathbb{R}^d} W_2(\delta_y, e^{-(t-s)L}\delta_y)\, d\mu_s(y). \end{aligned}$$

Our goal is to bound this by $O(\sqrt{t-s})$. In view of Proposition 6.14, it is sufficient to prove that for all $a > 0$,

$$W_2^2(\delta_y, e^{-(t-s)L}\delta_y) = O(t-s)\, O(e^{a|y|^2}).$$

This estimate is rather easy, since the left-hand side is just the variance of the solution of a linear diffusion equation, starting with a Dirac mass at y as initial datum. Without loss of generality, we assume $s = 0$, and write $\tilde{\mu}_t := e^{-tL}\delta_y$. For simplicity we write the computations in a sketchy way, but they are not hard to justify.

Since the initial datum is δ_y, its square exponential moment $\tilde{M}_\alpha$ of order α is $e^{\alpha|y|^2}$. With an argument similar to the proof of Proposition 6.14(ii), one can show that

$$0 \le t \le T \Longrightarrow \int e^{\alpha|x|^2}\, d\tilde{\mu}_t(x) \le C(T)(1 + \tilde{M}_\alpha) \le C(T)\, e^{\alpha|y|^2}.$$

Now, since $|\nabla V|(x) = O(e^{a|x|^2})$, $a < \alpha$, $|\nabla W * \mu_t|$ grows at most polynomially, and $\tilde{\mu}_t$ admits a square exponential moment of order α, we easily obtain

$$\frac{d}{dt}\int x\, d\tilde{\mu}_t = -\int \nabla(V + W*\mu_t)\, d\tilde{\mu}_t = \int O(e^{a|x|^2})\, d\tilde{\mu}_t = O(e^{\alpha|y|^2});$$

$$\frac{d}{dt}\int \frac{|x|^2}{2}\,d\tilde{\mu}_t = d - \int x\cdot\nabla(V+W*\mu_t)\,d\tilde{\mu}_t = O(e^{\alpha|y|^2}).$$

From these estimates we deduce that the time-derivative of the variance $V(\tilde{\mu}_t) := \int |x|^2\,d\tilde{\mu}_t - (\int x\,d\tilde{\mu}_t)^2$ is bounded by $O(e^{b|y|^2})$ for any $b > 0$. Since $\tilde{\mu}_0$ has zero variance, it follows that the variance of $\tilde{\mu}_t$ is $O(te^{b|y|^2})$, which was our goal. □

6.3.4 Regularity in phase space

Regularity estimates will be useful for Theorem 6.11. Equation (6.16) is a (weakly nonlinear) parabolic equation, for which regularization effects can be studied by standard tools. Some limits to the strength of the regularization are imposed by the regularity of V. So as not to be bothered by these nonessential considerations, we shall assume strong regularity conditions on V here. Then in Appendix 6.8 we shall prove the following estimates :

Proposition 6.19. *With the conventions of Subsection 6.3.1, assume in addition that V has all its derivatives growing at most polynomially at infinity. Then, for each $k \geq 0$ and for all $t_0 > 0$, $T > t_0$ there is a finite constant $C(t_0, T)$, only depending on t_0, T, k and a square exponential moment of the initial measure μ_0, such that the density f_t of μ_t is of class $\mathcal{C}^k$, with*

$$\sup_{t_0\leq t\leq T} \|f_t\|_{\mathcal{C}^k} \leq C(t_0, T).$$

If moreover $\beta > 0$, $\beta + \gamma > 0$, then $C(t_0, T)$ can be chosen to be independent of T for any fixed t_0.

Remark 6.20. For regular initial data and under some adequate assumptions on V and W, some regularity estimates on f_t/f_∞, where f_∞ is the limit density in large time, are established in [37, Lemma 6.7]. These estimates allow a much more precise uniform decay, but are limited to just one derivative. Here there will be no need for them.

6.3.5 Asymptotic behavior

In the "uniformly convex" case when $\beta + \gamma > 0$, the measure μ_t converges to a definite limit μ_∞ as $t \to \infty$. This was investigated in [74, 36, 37]. The following statement is a simple variant of [36, Theorems 2.1 and 5.1].

Proposition 6.21. *With the conventions of Subsection 6.3.1, assuming that $\beta > 0, \beta + 2\gamma > 0$, there exists a probability measure μ_∞ such that*

$$W_2(\mu_t, \mu_\infty) \leq Ce^{-\lambda t}, \qquad \lambda > 0.$$

Here the constants C and λ only depend on the initial datum μ_0.

6.4 The limit empirical measure

Consider the random time-dependent measure

$$\hat{\nu}_t^N := \frac{1}{N}\sum_{i=1}^N \delta_{Y_t^i}, \tag{6.59}$$

where $(Y_t^i)_{t\geq 0}$, $1 \leq i \leq N$, are N independent processes solving the same stochastic differential equation

$$dY_t^i = \sqrt{2}\, dB_t^i - [\nabla(V + W * \mu_t)](Y_t^i)\, dt,$$

and such that the law of Y_0^i is μ_0. As we already mentioned, for each t and i, Y_t^i is distributed according to the law μ_t. We call $\hat{\nu}_t^N$ the "limit empirical measure" because it is expected to be a rather accurate description, in some well-chosen sense, of the empirical measure $\hat{\mu}_t^N$ as $N \to \infty$.

Our estimates on μ_t, and the fact that $\hat{\nu}_t^N$ is the empirical measure for *independent* processes, are sufficient to imply good properties of concentration of $\hat{\nu}_t^N$ around its mean μ_t, as $N \to \infty$, for each t. But later on we shall use some estimates about the time-dependent measure (even to obtain a result of concentration for $\hat{\mu}_t^N$ with fixed t). To get such results, we shall study the time-regularity of $\hat{\nu}_t^N$. Our final goal in this section is the following

Proposition 6.22. *With the conventions of Subsection 6.3.1, for any $T \geq 0$ there are constants $C = C(T)$ and $a = a(T) > 0$ such that the limit empirical measure* (6.59) *satisfies*

$$\forall \Delta \in [0,T], \forall \varepsilon > 0, \qquad \mathbb{P}\Big[\sup_{t_0\leq s,t\leq t_0+\Delta} W_1(\hat{\nu}_s^N, \hat{\nu}_t^N) > \varepsilon\Big] \leq \exp\big(-N(a\,\varepsilon^2 - C\,\Delta)\big).$$

To prove Proposition 6.22, we shall use a bit of classical stochastic calculus tools.

6.4.1 SDE estimates

In this subsection we establish the following estimates of time regularity for the stochastic process Y_t : For all $T > 0$, there exist positive constants a and C such that, for all $s, t, t_0, \Delta \in [0,T]$,

$$\text{(i)} \qquad \mathbb{E}\,|Y_t - Y_s|^2 \leq C|t-s|$$

$$\text{(ii)} \qquad \mathbb{E}\,|Y_t - Y_s|^4 \leq C|t-s|^2$$

$$\text{(iii)} \quad \mathbb{E}\Big[\sup_{t_0\leq s\leq t\leq t_0+\Delta} \exp\big(a|Y_t - Y_s|^2\big)\Big] \leq 1 + C\,\Delta.$$

Proof. We start with (i). We use Itô's formula to write a stochastic equation on the process $(|Y_t - Y_s|^2)_{t\geq s}$:

$$|Y_t - Y_s|^2 = M_{s,t} + 2\,d\,(t-s) - 2\int_s^t (\nabla V(Y_u) + \nabla W * \mu_u(Y_u))\cdot(Y_u - Y_s)\,du,$$

where $M_{s,t}$, viewed as a process depending on t, is a martingale with zero expectation. Hence

$$\mathbb{E}\,|Y_t - Y_s|^2 = 2\,d\,(t-s) - 2\int_s^t \mathbb{E}\,(\nabla V(Y_u) + \nabla W * \mu_u(Y_u)) \cdot (Y_u - Y_s)\,du. \tag{6.60}$$

On one hand

$$\mathbb{E}\Big|(\nabla V(Y_u) + \nabla W * \mu_u(Y_u)) \cdot (Y_u - Y_s)\Big|^2 \le 4\Big(\mathbb{E}|\nabla V(Y_u)|^2 + \mathbb{E}|\nabla W * \mu_u(Y_u)|^2\Big) \Big(\mathbb{E}|Y_u|^2 + \mathbb{E}|Y_s|^2\Big). \tag{6.61}$$

On the other hand, by Proposition 6.14, μ_u has a finite square exponential moment, uniformly bounded for $u \in [0,T]$. More precisely, there exist $\alpha > 0$ and $M < +\infty$ such that $\displaystyle\int e^{\alpha|x|^2}\,d\mu_u(x) \le M$ for all $u \le T$. Since by assumption $|\nabla W(z)| \le L\,|z|$ and $|\nabla V(x)| = O(e^{\alpha|x|^2})$, we deduce

$$\sup_{s \le u \le T} \Big(\mathbb{E}\,(\nabla V(Y_u) + \nabla W * \mu_u(Y_u)) \cdot (Y_u - Y_s)\Big) < +\infty.$$

In view of (6.60), it follows that there exists a constant $C = C(T)$ such that

$$\mathbb{E}\,|Y_t - Y_s|^2 \le (2d + C)\,(t-s).$$

This concludes the proof of (i).

To establish (ii), we perform a very similar computation. For given s, let $Z_{s,t} := (|Y_t - Y_s|^4)_{t \ge s}$. Another application of Itô's formula yields

$$\mathbb{E}\,Z_{s,t} = 4(2+d)\int_s^t \mathbb{E}|Y_u - Y_s|^2\,du - 4\int_s^t \mathbb{E}|Y_u - Y_s|^2 (Y_u - Y_s) \cdot \big(\nabla V(Y_u) + \nabla W * \mu_u(Y_u)\big)\,du.$$

On one hand, from (i),

$$\int_s^t \mathbb{E}\,|Y_u - Y_s|^2\,du \le 2\,C\int_s^t (u-s)\,ds = C(t-s)^2.$$

On the other hand

$$\int_s^t \mathbb{E}\,|Y_u - Y_s|^2 (Y_u - Y_s) \cdot \big(\nabla V(Y_u) + \nabla W * \mu_u(Y_u)\big)\,du \le \left(\int_s^t \mathbb{E}\,Z_{s,u}\,du\right)^{3/4} \left(\int_s^t \mathbb{E}\,\big|\nabla V(Y_u) + \nabla W * \mu_u(Y_u)\big|^4\,du\right)^{1/4} \tag{6.62}$$

by Hölder's inequality. But again, since the measures μ_t admit a bounded square exponential moment, $\mathbb{E}\,|\nabla V(Y_u) + \nabla W * \mu_u(Y_u)|^4$ is bounded on $[0,T]$. We conclude that

$$\mathbb{E}\, Z_{s,t} \leq C \left((t-s)^2 + (t-s)^{1/4} \left(\int_s^t \mathbb{E}\, Z_{s,u}\, du \right)^{3/4} \right). \tag{6.63}$$

Then, with C standing again for various constants which are independent of s and t,

$$\mathbb{E}\, Z_{s,u} \leq C(\mathbb{E}|Y_u|^4 + \mathbb{E}|Y_s|^4) \leq 2\,C \sup_{0\leq u\leq T} \int |x|^4\, d\mu_u(x) \leq C;$$

so, from (6.63),

$$\mathbb{E}\, Z_{s,t} \leq C((t-s)^2 + (t-s)^{1/4}(t-s)^{3/4}) \leq C(t-s),$$

and by (6.63) again we successively obtain

$$\mathbb{E}\, Z_{s,t} \leq C(t-s)^{7/4},$$

and finally

$$\mathbb{E}\, Z_{s,t} \leq C(t-s)^2.$$

This concludes the proof of (ii).

We finally turn to the proof of (iii). Without real loss of generality, we set $t_0 = 0$. We shall proceed as in the proof of Proposition 6.14, and prove the existence of some constant C and some continuous positive function a on $\mathbb{R}_+$ such that

$$\mathbb{E} \left(\sup_{0\leq s\leq t\leq \Delta\leq T} \exp\left(a(t)|Y_t - Y_s|^2\right) \right) \leq 1 + C\,\Delta. \tag{6.64}$$

Let $a(t)$ be a smooth function, and

$$Z_{s,t} := e^{a(t)|Y_t - Y_s|^2}.$$

By Itô's formula,

$$Z_{s,t} = 1 + M_{s,t}$$
$$+ \int_s^t \Big[2a(u)\Big(d + 2a|Y_u - Y_s|^2 - (\nabla V + \nabla W * \mu_u)(Y_u)\cdot(Y_u - Y_s)\Big) + a'(u)|Y_u - Y_s|^2 \Big]\, Z_{s,u}\, du$$

where

$$M_{s,t} := \int_s^t a(u)\,(Y_u - Y_s)\, Z_u\, dB_u.$$

For each s, $M_{s,t}$, viewed as a stochastic process in t, is a martingale.

By Young's inequality, for any $b > 0$,

$$-2\big(\nabla V + \nabla W * \mu_u\big)(Y_u) \cdot (Y_u - Y_s) \le b\big|Y_u - Y_s\big|^2 + \frac{1}{b}\big|\nabla V + \nabla W * \mu_u\big|^2(Y_u).$$

So, by letting

$$A_u := a(u) \left[2\,d + \frac{1}{b}\big|\nabla V(Y_u) + \nabla W * \mu_u(Y_u)\big|^2\right]$$

and

$$B(u) := a'(u) + 4\,a^2(u) + b\,a(u)$$

we obtain

$$Z_{s,t} \le 1 + M_{s,t} + \int_s^t [A_u + B(u)|Y_u - Y_s|^2]\, Z_{s,u}\, du.$$

We choose a in such a way that the function B is identically zero, that is

$$a(u) = e^{-b\,u} \left(\frac{1}{a(0)} + 4\,\frac{1 - e^{-b\,u}}{b}\right)^{-1},$$

where $a(0)$ is to be fixed later. Then

$$Z_{s,t} \le 1 + M_{s,t} + \int_s^t A_u\, Z_{s,u}\, du$$

from which it is clear that

$$\mathbb{E} \sup_{s \le t \le \Delta} Z_{s,t} \le 1 + \mathbb{E} \sup_{s \le t \le \Delta} M_{s,t} + \int_s^\Delta \mathbb{E}\, A_u\, Z_{s,u}\, du. \tag{6.65}$$

By Cauchy-Schwarz and Doob's inequalities,

$$\left(\mathbb{E} \sup_{s \le t \le \Delta} M_{s,t}\right)^2 \le \mathbb{E} \left|\sup_{s \le t \le \Delta} M_{s,t}\right|^2 \le 2 \sup_{s \le t \le \Delta} \mathbb{E}\, |M_{s,t}|^2. \tag{6.66}$$

Also, by Itô's formula and the Cauchy-Schwarz inequality again,

$$\begin{aligned}\mathbb{E}\, |M_{s,t}|^2 &= \int_s^t a(u)^2\, \mathbb{E}|Y_u - Y_s|^2\, Z_{s,u}^2\, du \\ &\le \frac{1}{2} \int_s^t a(u)^2 \left(\mathbb{E}\, |Y_u - Y_s|^4\right)^{1/2} \left(\mathbb{E}\, Z_{s,u}^4\right)^{1/2} du. \end{aligned} \tag{6.67}$$

In view of (ii), there exists a constant C such that

$$\mathbb{E}\, |Y_u - Y_s|^4 \le C\, (u - s)^2. \tag{6.68}$$

Furthermore,

$$\mathbb{E}\, Z_{s,u}^4 = \mathbb{E} \exp 4\, a(u)|Y_u - Y_s|^2 \le \left(\mathbb{E} \exp 16\, a(u)|Y_u|^2\right)^{1/2} \left(\mathbb{E} \exp 16\, a(u)|Y_s|^2\right)^{1/2}. \tag{6.69}$$

Recall from Proposition 6.14 that there exist constants M and $\alpha > 0$ such that

$$\sup_{s \le u \le \Delta} \int e^{\alpha |y|^2} \, d\mu_u(y) \le M.$$

If we choose $a(0) \le \alpha/16$, the decreasing property of a will ensure that $a(u) \le \alpha/16$ for all $u \in [0, \Delta]$, and

$$\mathbb{E} \exp 16\, a(u) |Y_u|^2 \left(= \int e^{16\, a(u)|y|^2} \, d\mu_u(y) \right) \le M.$$

Then, from (6.69),

$$\sup_{s \le u \le \Delta} \mathbb{E}\, Z_{s,u}^4 \le M.$$

Now, from (6.67) and (6.68) we deduce

$$\sup_{s \le t \le \Delta} \mathbb{E}\, |M_{s,t}|^2 \le C\,(t-s)^2.$$

Combining this with (6.66), we conclude that

$$\mathbb{E} \sup_{s \le t \le \Delta} M_{s,t} \le C\, \Delta.$$

In the same way, we can prove that $\mathbb{E}\,(A_t Z_{s,t})$ is bounded for $t \in [s, \Delta]$ by bounding $\mathbb{E}\, Z_{s,t}^2$ and $\mathbb{E}\, A_t^2$. This concludes the proof of (6.64), and therefore of (iii) above. □

6.4.2 Time-regularity of the limit empirical measure

We are now ready to prove Proposition 6.22.

On one hand

$$W_1(\hat{\nu}_s^N, \hat{\nu}_t^N) \le \frac{1}{N} \sum_{i=1}^N |Y_t^i - Y_s^i|,$$

so

$$\mathbb{P}\left[\sup_{0 \le s \le t \le \Delta} W_1(\hat{\nu}_s^N, \hat{\nu}_t^N) > \varepsilon \right] \le \mathbb{P}\left[\frac{1}{N} \sum_{i=1}^N V^i > \varepsilon \right] \tag{6.70}$$

where

$$V^i := \sup_{0 \le s \le t \le \Delta} |Y_t^i - Y_s^i|.$$

By Chebyshev's exponential inequality and the independence of the $(Y_t^i - Y_s^i)$,

$$\mathbb{P}\left[\frac{1}{N} \sum_{i=1}^N V^i > \varepsilon \right] \le \exp\left(-N \sup_{\zeta \ge 0} \left[\varepsilon \zeta - \ln \mathbb{E} \exp(\zeta V^1) \right] \right).$$

But, for any given ζ and $\omega \ge 0$,

$$\mathbb{E} \exp(\zeta V^1) \le \mathbb{E} \exp\left(\zeta \Big(\frac{\omega^2 + (V^1)^2}{2\,\omega} \Big) \right) \le \exp \frac{\zeta\, \omega}{2}\, \mathbb{E} \exp \frac{\zeta}{2\,\omega} (V^1)^2.$$

Let $\omega = \dfrac{\zeta}{2\,a}$, so that $\dfrac{\zeta}{2\,\omega} = a$. Then, from estimate (iii) in Subsection 6.4.1,

$$\mathbb{E}\exp\frac{\zeta}{2\,\omega}(V^1)^2 \leq 1 + C\,\Delta,$$

uniformly in s and Δ. Hence, for any $\zeta > 0$,

$$\mathbb{E}\exp(\zeta V^1) \leq \mathbb{E}\exp\frac{\zeta^2}{4\,a}\,(1 + C\,\Delta).$$

Consequently,

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} V^i > \varepsilon\right] &\leq \exp\left(-N\sup_{\zeta\geq 0}\left[\varepsilon\zeta - \frac{\zeta^2}{4} - \ln(1 + C\,\Delta)\right]\right)\\
&= \exp\Big(-N\left[a\,\varepsilon^2 - \ln(1 + C\,\Delta)\right]\Big)\\
&\leq \exp\left(-N\left[a\,\varepsilon^2 - C\,\Delta\right]\right).
\end{aligned}$$

The proof of Proposition 6.22 follows by (6.70). □

6.5 Coupling

We now (as is classical) reduce the proof of convergence for $\hat{\mu}_t^N$ to a proof of convergence for the empirical measure $\hat{\nu}_t^N$ constructed on the auxiliary independent system (Y_t^i). The final goal of this section is the following estimate.

Proposition 6.23. *With the conventions of Subsection 6.3.1,*

$$W_1(\hat{\mu}_t^N, \mu_t) \leq \Gamma \int_0^t e^{-\alpha(t-s)} W_1(\hat{\nu}_s^N, \mu_s)\,ds + W_1(\hat{\nu}_t^N, \mu_t),$$

where Γ *is defined by* (6.51), *and* $\alpha := \beta + 2\min(\gamma, 0)$.

Proof. For the sake of simplicity we give a slightly sketchy proof. We couple the stochastic systems (X_t^i) and (Y_t^i) by assuming that (i) $X_0^i = Y_0^i$ and (ii) both systems are driven by the *same* Brownian processes B_t^i. In particular, for each $i \in \{1, \ldots, N\}$, the process $X_t^i - Y_t^i$ satisfies the equation

$$d(X_t^i - Y_t^i) = -\big(\nabla V(X_t^i) - \nabla V(Y_t^i)\big)\,dt - \Big(\nabla W * \hat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i)\Big)\,dt. \qquad (6.71)$$

From (6.71) we deduce

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}|X_t^i - Y_t^i|^2 = &-(\nabla V(X_t^i) - \nabla V(Y_t^i)) \cdot (X_t^i - Y_t^i)\\
&- \big(\nabla W * \hat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i)\big) \cdot (X_t^i - Y_t^i). \qquad (6.72)
\end{aligned}$$

Our convexity assumption on V implies

$$-(\nabla V(X_t^i) - \nabla V(Y_t^i)) \cdot (X_t^i - Y_t^i) \leq -\beta |X_t^i - Y_t^i|^2;$$

so the main issue consists in the treatment of the quantity $\nabla W * \hat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i)$ appearing in the right-hand side of (6.72). There are (at least) two options here. The first one consists in writing

$$\nabla W * \hat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i) = \\ (\nabla W * \hat{\mu}_t^N - \nabla W * \mu_t)(X_t^i) + \big(\nabla W * \mu_t(X_t^i) - \nabla W * \mu_t(Y_t^i)\big); \quad (6.73)$$

while the second one consists in forcing the introduction of $\hat{\nu}_t^N$ as follows :

$$\nabla W * \hat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i) = \\ \frac{1}{N}\sum_{j=1}^N \big[\nabla W(X_t^i - X_t^j) - \nabla W(Y_t^i - Y_t^j)\big] - (\nabla W * \hat{\nu}_t^N - \nabla W * \mu_t)(Y_t^i). \quad (6.74)$$

Both options are interesting and lead to slightly different computations. Since both lines of computations might be useful in other contexts, we shall sketch them one after the other. The second option leads to better bounds, but at the price of more complications (in particular, we shall need to sum over the index i at an early stage).

First option : We start as in (6.73). In view of our assumption on D^2W, the Lipschitz norm of $\nabla W(X_t^i - \cdot)$ is bounded by Γ. Therefore, by the Kantorovich-Rubinstein dual formulation (6.8),

$$\Big|\nabla W * (\hat{\mu}_t^N - \mu_t)(X_t^i)\Big| = \left|\int_{\mathbb{R}^d} \nabla W(X_t^i - y)\, d(\hat{\mu}_t^N - \mu_t)(y)\right| \leq \Gamma\, W_1(\hat{\mu}_t^N, \mu_t),$$

and then our assumptions on V and W imply

$$\frac{1}{2}\frac{d}{dt}|X_t^i - Y_t^i|^2 \leq -(\gamma+\beta)\,|X_t^i - Y_t^i|^2 + \Gamma\, W_1(\hat{\mu}_t^N, \mu_t)\,|X_t^i - Y_t^i|.$$

In other words, $|X_t^i - Y_t^i|$ satisfies the differential inequality

$$\frac{d}{dt}|X_t^i - Y_t^i| + (\beta+\gamma)\,|X_t^i - Y_t^i| \leq \Gamma\, W_1(\hat{\mu}_t^N, \mu_t)$$

(X_t^i and Y_t^i separately are not Lipschitz functions of t, but their difference is). Hence, by Gronwall's lemma,

$$|X_t^i - Y_t^i| \leq \Gamma \int_0^t e^{-(\beta+\gamma)(t-s)}\, W_1(\hat{\mu}_s^N, \mu_s)\, ds.$$

Now we sum over i ; by convexity of the distance W_1 and triangular inequality, we obtain

$$\begin{aligned} W_1(\hat{\mu}_t^N, \hat{\nu}_t^N) &\leq \frac{1}{N}\sum_{i=1}^N |X_t^i - Y_t^i| \leq \Gamma \int_0^t e^{-(\beta+\gamma)(t-s)}\, W_1(\hat{\mu}_s^N, \mu_s)\, ds \\ &\leq \Gamma \int_0^t e^{-(\beta+\gamma)(t-s)}\, \big[W_1(\hat{\mu}_s^N, \hat{\nu}_s^N) + W_1(\hat{\nu}_s^N, \mu_s)\big]\, ds. \end{aligned}$$

By using Gronwall's lemma again, we deduce

$$W_1(\hat{\mu}_t^N, \hat{\nu}_t^N) \leq \Gamma \int_0^t e^{-(\beta+\gamma-\Gamma)(t-s)}\, W_1(\hat{\nu}_s^N, \mu_s)\, ds.$$

By applying the triangular inequality for W_1, we conclude to the validity of Proposition (6.23), only with α replaced by the (a priori smaller) quantity $\beta + \gamma - \Gamma$.

Second option : Now we start with (6.74). This time we sum over i right from the beginning :

$$\frac{1}{2}\frac{d}{dt}\sum_{i=1}^N |X_t^i - Y_t^i|^2 = -\sum_{i=1}^N (\nabla V(X_t^i) - \nabla V(Y_t^i)) \cdot (X_t^i - Y_t^i) - \frac{1}{N}\sum_{i,j=1}^N (A_t^{ij} + B_t^{ij})$$

where

$$A_t^{ij} = (\nabla W(X_t^i - X_t^j) - \nabla W(Y_t^i - Y_t^j))\,\cdot\,(X_t^i - Y_t^i)$$

and

$$B_t^{ij} = (W(Y_t^i - Y_t^j) - \nabla W * \mu_t(Y_t^i))\,\cdot\,(X_t^i - Y_t^i).$$

Since ∇W is an odd function and $D^2W(x) \geq \gamma I$ for all $x \in \mathbb{R}^d$, we have

$$\begin{aligned} A_t^{ij} + A_t^{ji} &= \big(\nabla W(X_t^i - X_t^j) - \nabla W(Y_t^i - Y_t^j)\big)\,\cdot\,\big((X_t^i - X_t^j) - (Y_t^i - Y_t^j)\big) \\ &\geq \gamma\big|(X_t^i - X_t^j) - (Y_t^i - Y_t^j)\big|^2, \end{aligned}$$

whence

$$-\sum_{i,j=1}^N A_t^{ij} \leq -\frac{\gamma}{2}\sum_{i,j=1}^N \big|(X_t^i - X_t^j) - (Y_t^i - Y_t^j)\big|^2 \leq -2N\gamma^- \sum_{i=1}^N |X_t^i - Y_t^i|^2$$

where $\gamma^- = \min(\gamma, 0)$.

Then

$$-\sum_{j=1}^N B_t^{ij} = -(X_t^i - Y_t^i)\cdot(\nabla W * \hat{\nu}_t^N(Y_t^i) - \nabla W * \mu_t(Y_t^i)).$$

Our assumption on D^2W implies that the Lipschitz norm of $\nabla W(Y_t^i - \cdot\,)$ is bounded by Γ ; so, by the Kantorovich-Rubinstein dual formulation (6.8),

$$\Big|\nabla W * (\hat{\nu}_t^N - \mu_t)(Y_t^i)\Big| = \left|\int_{\mathbb{R}^d} \nabla W(Y_t^i - y)\, d(\hat{\nu}_t^N - \mu_t)(y)\right| \leq \Gamma\, W_1(\hat{\nu}_t^N, \mu_t).$$

Collecting all terms we finally obtain

$$\frac{1}{2}\frac{d}{dt}\sum_{i=1}^N |X_t^i - Y_t^i|^2 \leq -(\beta + 2\gamma^-)\sum_{i=1}^N |X_t^i - Y_t^i|^2 + \Gamma\sum_{i=1}^N |X_t^i - Y_t^i| W_1(\hat{\nu}_t^N, \mu_t).$$

Then, since $\sum_{i=1}^{N} |X_t^i - Y_t^i| \leq \left(N \sum_{i=1}^{N} |X_t^i - Y_t^i|^2\right)^{1/2}$, the function $y(t) := \left(\frac{1}{N}\sum_{i=1}^{N} |X_t^i - Y_t^i|^2\right)^{1/2}$ satisfies the differential inequality

$$y'(t) + (\beta + 2\gamma^-)y(t) \leq \Gamma\, W_1(\hat{\nu}_t^N, \mu_t),$$

so that

$$\left(\frac{1}{N}\sum_{i=1}^{N} |X_t^i - Y_t^i|^2\right)^{1/2} \leq \Gamma \int_0^t e^{-(\beta+2\gamma^-)(t-s)}\, W_1(\hat{\nu}_s^N, \mu_s)\, ds.$$

The conclusion follows by triangular inequality again since

$$W_1(\hat{\mu}_t^N, \hat{\nu}_t^N) \leq W_2(\hat{\mu}_t^N, \hat{\nu}_t^N) \leq \left(\frac{1}{N}\sum_{i=1}^{N} |X_t^i - Y_t^i|^2\right)^{1/2}.$$

□

Remark 6.24. Not only does the "second option" in the proof lead to better bounds, it also provides an estimate of the distance between $\hat{\mu}$ and $\hat{\nu}$ in the W_2 distance, which is stronger than the W_1 distance. However, we do not take any advantage of this refinement.

6.6 Conclusion

In this section, we paste together all the estimates established in the previous sections, so as to prove Theorems 6.7 to 6.11.

6.6.1 Concentration estimates

We start with the proof of Theorem 6.7. By C we shall denote various constants depending on T, on our assumptions on V and W, and also on $\int e^{\alpha|x|^2}\, d\mu_0(x)$, for some $\alpha > 0$.

From Proposition 6.23,

$$\sup_{0\leq t\leq T} W_1(\hat{\mu}_t^N, \mu_t) \leq (\Gamma e^{|\alpha|T} + 1) \sup_{0\leq t\leq T} W_1(\hat{\nu}_t^N, \mu_t)\, ds.$$

In particular, there is a constant C such that

$$\mathbb{P}\left[\sup_{0\leq t\leq T} W_1(\hat{\mu}_t^N, \mu_t) > \varepsilon\right] \leq \mathbb{P}\left[\sup_{0\leq t\leq T} W_1(\hat{\nu}_t^N, \mu_t) > \tilde{\varepsilon}\right], \qquad \tilde{\varepsilon} = \frac{\varepsilon}{C}. \tag{6.75}$$

From Corollary 6.15 and Theorem 6.1 we know that

$$\sup_{0\leq t\leq T} \mathbb{P}\left[W_1(\hat{\nu}_t^N, \mu_t) > \tilde{\varepsilon}\right] \leq e^{-K\, N\, \tilde{\varepsilon}^2}$$

for all $t \in [0,T]$, $N \geq N_0 \max(\tilde{\varepsilon}^{-(d'+2)}, 1)$ $(d' > d)$. The issue now is to "exchange" sup and $\mathbb{P}$ in this estimate. As we shall see, this is authorized by the continuity estimates on $\hat{\nu}_t^N$ and μ_t.

Let $\Delta > 0$ (to be fixed later on), and let M be the integer part of $T/\Delta + 1$. We decompose the interval $[0,T]$ as

$$[0,T] = [0,\Delta] \cup [\Delta, 2\Delta] \cup \ldots \cup [(M-1)\Delta, T] \subset \bigcup_{h=0}^{M-1} [h\Delta, (h+1)\Delta].$$

Proposition 6.16 guarantees that, if $\Delta \leq a\tilde{\varepsilon}^2$ for some a small enough, then

$$h\Delta \leq t \leq (h+1)\Delta \Longrightarrow W_1(\mu_t, \mu_{h\Delta}) \leq \frac{\tilde{\varepsilon}}{2}. \tag{6.76}$$

Then, by triangular inequality and (6.76),

$$\mathbb{P}\left[\sup_{0\leq t\leq T} W_1(\hat{\nu}_t^N, \mu_t) > \tilde{\varepsilon}\right]$$

$$\leq \mathbb{P}\left[\sup_{h=0,\ldots,M-1} \sup_{h\Delta\leq t\leq (h+1)\Delta} W_1(\hat{\nu}_t^N, \mu_t) > \tilde{\varepsilon}\right]$$

$$\leq \mathbb{P}\Bigg[\sup_{h=0,\ldots,M-1} \sup_{h\Delta\leq t\leq (h+1)\Delta} W_1(\hat{\nu}_t^N, \nu_{h\Delta}^N) + \sup_{h=0,\ldots,M-1} W_1(\hat{\nu}_{h\Delta}^N, \mu_{h\Delta})$$
$$+ \sup_{h=0,\ldots,M-1} \sup_{h\Delta\leq t\leq (h+1)\Delta} W_1(\mu_{h\Delta}, \mu_t) > \tilde{\varepsilon})\Bigg]$$

$$\leq \mathbb{P}\left[\sup_{h=0,\ldots,M-1} \sup_{h\Delta\leq t\leq (h+1)\Delta} W_1(\hat{\nu}_t^N, \hat{\nu}_{h\Delta}^N) + \sup_{h=0,\ldots,M-1} W_1(\hat{\nu}_{h\Delta}^N, \mu_{h\Delta}) > \frac{\tilde{\varepsilon}}{2}\right],$$

which can be bounded by

$$\mathbb{P}\left[\sup_{h=0,\ldots,M-1} \sup_{h\Delta\leq t\leq (h+1)\Delta} W_1(\hat{\nu}_t^N, \hat{\nu}_{h\Delta}^N) > \frac{\tilde{\varepsilon}}{4}\right] + \mathbb{P}\left[\sup_{h=0,\ldots,M-1} W_1(\hat{\nu}_{h\Delta}^N, \mu_{h\Delta}) > \frac{\tilde{\varepsilon}}{4}\right].$$

By Corollary 6.15 and Theorem 6.1, there exist some constants C and N_0 such that

$$\mathbb{P}\left[W_1(\hat{\nu}_{h\Delta}^N, \mu_{h\Delta}) \geq \frac{\tilde{\varepsilon}}{4}\right] \leq \exp(-C\,N\,\tilde{\varepsilon}^2)$$

for all $h = 0, \ldots, M-1$, and $N \geq N_0 \max(\tilde{\varepsilon}^{-(d'+2)}, 1)$. Hence

$$\mathbb{P}\left[\sup_{h=0,\ldots,M-1} W_1(\hat{\nu}_{h\Delta}^N, \mu_{h\Delta}) > \frac{\tilde{\varepsilon}}{4}\right] \leq \sum_{h=0}^{M-1} \mathbb{P}\left[W_1(\hat{\nu}_{h\Delta}^N, \mu_{h\Delta}) > \frac{\tilde{\varepsilon}}{4}\right] \leq M \exp(-C\,N\,\tilde{\varepsilon}^2). \tag{6.77}$$

On the other hand, from Proposition 6.22 we deduce

$$\mathbb{P}\left[\sup_{h\Delta\le t\le (h+1)\Delta} W_1(\hat\nu_t^N,\hat\nu_{h\Delta}^N) > \frac{\tilde\varepsilon}{4}\right] \le \exp\Big(-N\big(\frac{a}{4}\,\tilde\varepsilon^2 - C\,\Delta\big)\Big)$$

for all $h=0,\dots,M-1$ and $\tilde\varepsilon>0$, so

$$\mathbb{P}\left[\sup_{h=0,\dots,M-1}\ \sup_{h\Delta\le t\le (h+1)\Delta} W_1(\hat\nu_t^N,\hat\nu_{h\Delta}^N) > \frac{\tilde\varepsilon}{4}\right] \le M\exp\Big(-N\big(\frac{a}{4}\,\tilde\varepsilon^2 - C\,\Delta\big)\Big). \qquad (6.78)$$

We can assume that $\Delta\le \frac{a}{8C}\tilde\varepsilon^2$, and $M\le CT/\tilde\varepsilon^2+1$; then we can bound the right-hand side of (6.78) by

$$M\exp\Big(-\frac{a}{8}\,N\,\tilde\varepsilon^2\Big) \le C\left(1+\frac{T}{\tilde\varepsilon^2}\right)\ \exp\Big(-\frac{a}{8}\,N\,\tilde\varepsilon^2\Big) \qquad (6.79)$$

From (6.77) and (6.79) we deduce that, for Δ small enough (depending on ε !),

$$\mathbb{P}\left[\sup_{0\le t\le T} W_1(\hat\nu_t^N,\mu_t) > \tilde\varepsilon\right] \le 2\,C\left(1+\frac{T}{\tilde\varepsilon^2}\right)\ \exp(-K\,N\,\tilde\varepsilon^2) \qquad (6.80)$$

for $N\ge N_0\max(\tilde\varepsilon^{-(d'+2)},1)$. So we deduce from (6.80) that

$$\mathbb{P}\left[\sup_{0\le t\le T} W_1(\hat\nu_t^N,\mu_t) > \tilde\varepsilon\right] \le \exp\left(\ln\left(C\left(\frac{T}{\tilde\varepsilon^2}+1\right)\right) - K\,N\,\tilde\varepsilon^2\right),$$

where again C,K stand for various positive constants, and $N\ge \max(N_0\,\varepsilon^{-(d'+2)},1)$. This concludes the proof of Theorem 6.7. □

6.6.2 Uniform in time estimates

Now, we shall focus on the case when $\beta>0, \beta+2\,\gamma>0$, and derive Theorem 6.9 by a slightly refined estimate.

Let us start again from the bound

$$W_1(\hat\mu_t^N,\mu_t)\le \Gamma\int_0^t e^{-\alpha(t-s)}W_1(\hat\nu_s^N,\mu_s)\,ds + W_1(\hat\nu_t^N,\mu_t)$$

where $\alpha:=\beta+2\min(\gamma,0)$ is positive. Let $\Delta>0$ (to be fixed later on), and k be the integer part of t/Δ. If $W_1(\hat\mu_t^N,\mu_t)$ is larger than ε, then

$$\begin{cases} \text{either } W_1(\hat\nu_t^N,\mu_t)\ge \dfrac{\varepsilon}{2} \\[2ex] \text{or } \exists j\in\{0,\dots,k\};\quad \displaystyle\int_{j\Delta}^{(j+1)\Delta} e^{-\alpha(t-s)}W_1(\hat\nu_s^N,\mu_s)\,ds \ge \frac{\varepsilon}{2^{k+2-j}\Gamma}. \end{cases}$$

Indeed, $(\varepsilon/2) + \sum_{j\le k}(\varepsilon/2^{k+2-j}) \le \varepsilon$. As a consequence,

$$\begin{cases} \text{either } W_1(\hat{\nu}_t^N, \mu_t) > \dfrac{\varepsilon}{2} \\ \\ \text{or } \exists j \in \{0,\ldots,k\}; \qquad \sup\limits_{j\Delta\le s\le (j+1)\Delta} W_1(\hat{\nu}_s^N, \mu_s) > \dfrac{\varepsilon\alpha e^{\alpha[t-(j+1)\Delta]}}{2^{k+2-j}\Gamma}. \end{cases}$$

Since, for $t \in [j\Delta, (j+1)\Delta]$,

$$\frac{e^{\alpha[t-(j+1)\Delta]}}{2^{k+2-j}} \ge \frac{e^{\alpha(k-j-1)\Delta}}{2^{k-j+2}} = \left(\frac{1}{4e^{\alpha\Delta}}\right)\left(\frac{e^{\alpha\Delta}}{2}\right)^{k-j},$$

we conclude to the existence of a constant C such that

$$\mathbb{P}\left[W_1(\hat{\mu}_t^N, \mu_t) > \varepsilon\right] \le \mathbb{P}\left[W_1(\hat{\nu}_t^N, \mu_t) > \frac{\varepsilon}{2}\right] + \sum_{j=0}^{k} \mathbb{P}\left[\sup_{j\Delta\le s\le(j+1)\Delta} W_1(\hat{\nu}_s^N, \mu_s) > C\varepsilon\left(\frac{e^{\alpha\Delta}}{2}\right)^{k-j}\right]. \tag{6.81}$$

We already know that the first term in the right-hand side in (6.81) is bounded by $e^{-\lambda N\varepsilon^2}$ for some constant $\lambda > 0$, and so we focus on the other terms.

In the proof of Theorem 6.7, we have established that there are constant C and λ, depending on Δ and on bounds on square exponential moments for μ_0, such that

$$\mathbb{P}\left[\sup_{0\le s\le\Delta} W_1(\hat{\nu}_s^N, \mu_s) > \delta\right] \le C\left(1 + \frac{\Delta}{\delta^2}\right) e^{-\lambda N\delta^2}. \tag{6.82}$$

Proposition 6.14 guarantees that these square exponential bounds also hold true for μ_t, uniformly in t. Thus we can apply (6.82) with $\mu_{j\Delta}$ taken as initial datum, and get

$$\mathbb{P}\left[\sup_{j\Delta\le s\le(j+1)\Delta} W_1(\hat{\nu}_s^N, \mu_s) > \delta\right] \le Ce^{-\lambda N\delta^2}, \tag{6.83}$$

as soon as $N \ge N_0 \max(\delta^{-(d'+2)}, 1)$.

We now use (6.83) to bound the sum appearing in the right-hand side of (6.81). Choose Δ large enough that

$$\theta := \frac{e^{\alpha\Delta}}{2} > 1.$$

Applying (6.83) with δ replaced by $C\theta^{j-k}\varepsilon$, we can bound the sum in the right-hand side of (6.81) by

$$C\sum_{j=0}^{k} \exp\bigl(-K\theta^{2(k-j)}N\varepsilon^2\bigr)$$

for $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$, where C, K and N_0 are again positive constants. Since again θ is larger than 1, there is a constant $a > 0$ such that $\theta^{2(k-j)} \geq a(k-j)$, so the sum above is bounded by

$$C\left(e^{-KN\varepsilon^2} + \sum_{\ell=1}^{\infty} e^{-K\ell N\varepsilon^2}\right) \leq C\left(e^{-KN\varepsilon^2} + \frac{e^{-KN\varepsilon^2}}{1 - e^{-KN\varepsilon^2}}\right).$$

If N_0 is large enough, our assumption $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$ implies that $e^{-KN\varepsilon^2}$ is always less than $1/2$, so that the above sum can be bounded by just $Ce^{-KN\varepsilon^2}$. This concludes the proof of the first point of Theorem 6.9.

The second point is proved by writing

$$W_1(\hat{\mu}_t^N, \mu_\infty) \leq W_1(\hat{\mu}_t^N, \mu_t) + W_1(\mu_t, \mu_\infty) \leq W_1(\hat{\mu}_t^N, \mu_t) + Ce^{-\lambda t}$$

successively by the triangular inequality for Wasserstein distance and use of Proposition 6.21. Then the result follows from the uniform estimate obtained above. □

6.6.3 Data reconstruction

We finally consider Theorem 6.11. Proposition 6.19 ensures that, as $t \to \infty$, f_t is uniformly bounded in C^k, where k is arbitrarily large. Since f_t converges to f_∞ as $t \to \infty$, we deduce that f_∞ is Lipschitz. Then Theorem 6.9 and Proposition 6.12 together imply Theorem 6.11. □

6.7 Appendix : metric entropy of a probability space

We now prove the covering result used in Section 6.2.1, as a particular case of a more general estimate. Let E be a Polish space, we look for an upper bound on the number $\mathcal{N}_p(E, \delta) := m(\mathcal{P}(E), \delta)$ of balls of radius δ in Wasserstein distance W_p needed to cover the space $\mathcal{P}(E)$ of probability measures on E. We use the same strategy as in [43, Exercise 6.2.19], where the Lévy distance is used instead of the Wasserstein distance.

Theorem 6.25. *Let (E, d) be a Polish space with finite diameter D. For any $r > 0$, define $N(E, r)$ as the minimal number of balls needed to cover E by balls of radius r. Then, for all $p \geq 1$ and $\delta \in (0, D)$, the space $\mathcal{P}(E)$ can be covered by $\mathcal{N}_p(E, \delta)$ balls of radius δ in W_p distance, with*

$$\mathcal{N}_p(E, \delta) \leq \left(\frac{8eD}{\delta}\right)^{pN\left(E, \frac{\delta}{2}\right)}. \tag{6.84}$$

Remark 6.26. The W_p distance between any two probability measures on E is at most D, so, for all $\delta \geq D$, we have the trivial estimate $\mathcal{N}_p(E, \delta) = 1$.

Proof. Let $r > 0$, and let $\{x_j\}_{1\le j\le N(E,r)}$ be such that E is covered by the balls $B(x_j, r)$ with centers $x_j \in E$ and radius r. For simplicity we shall write $N = N(E, r)$.

In a **first step** we prove that for any $\mu \in \mathcal{P}(E)$ there exist nonnegative real numbers $(\beta_j)_{1\le j\le N}$, with $\sum_{j=1}^{N} \beta_j = 1$, such that

$$W_p(\mu, \tilde{\mu}) \le r, \qquad \tilde{\mu} := \sum_{j=1}^{N} \beta_j \delta_{x_j}.$$

For this we first replace the balls $B(x_j, r)$'s by the sets $\tilde{B}_j$'s defined by

$$\forall j, \qquad \tilde{B}_j = B(x_j, r) \setminus \bigcup_{k\le j-1} B(x_k, r),$$

so that E is partitioned into the $\tilde{B}_j$'s. Next define

$$\beta_j = \mu[\tilde{B}_j].$$

It is easy to check that the required properties are fulfilled. Indeed, we may transport μ onto $\tilde{\mu} = \sum_{j=1}^{N} \beta_j \delta_{x_j}$ by sending all x's in $\tilde{B}_j$ onto x_j, for each $j = 1, \ldots, N$: the cost of this transport is bounded by $\sum_{j=1}^{N} r^p \mu(\tilde{B}_j) = r^p$.

In the **second step** we introduce an integer K (whose value will be made more precise later on), and consider the set

$$\mathcal{C}_K := \left\{ \sum_{j=1}^{N} \alpha_j \delta_{x_j}; \quad (\alpha_j)_{1\le j\le N} \in \mathcal{A}_K \right\} \subset \mathcal{P}(E),$$

where $\mathcal{A}_K$ is the set of all N-tuples $(\alpha_j)_{1\le j\le N}$, such that each α_j is of the form k_j/K, $k_j \in \mathbb{N}$, and $\sum_{j=1}^{N} \alpha_j = 1$.

Given a probability measure $\tilde{\mu} = \sum_{i=1}^{N} \beta_i \delta_{x_i}$ (where $(\beta_i)_i$ does not necessarily belong to $\mathcal{A}_K$), there exists μ' in $\mathcal{C}_K$ such that

$$W_p(\mu', \tilde{\mu}) \le D \left(\frac{N}{K}\right)^{1/p}. \tag{6.85}$$

To prove (6.85), we define n_j as the integer part $[K\beta_j]$ of $K\beta_j$ and J as the first integer such that

$$\sum_{j=1}^{J} (n_j + 1) + \sum_{j=J+1}^{N} n_j = K.$$

Since $\sum_{j=1}^N \beta_j = 1$, it is clear that $J \leq N$. Then we define a measure $\mu' \in \mathcal{C}_K$ by $\mu' = \sum_{j=1}^N \alpha_j \delta_{x_j}$, where

$$\alpha_j = \begin{cases} \frac{n_j+1}{K} & \text{for } j = 1, \ldots, J \\ \frac{n_j}{K} & \text{for } j = J+1, \ldots, N. \end{cases}$$

Let us bound the distance between μ and μ'. For that we gradually define a transport plan between $\tilde{\mu}$ and μ' in the following way : first of all, at each point x_i, the mass n_i/K stays in place. Then, the remaining masses $\beta_i - n_i/K$ are redistributed as follows : all the remaining mass at $x_1, \ldots, x_\ell$ is brought to x_1, together with possibly a bit of mass at $x_{\ell+1}$, until a total mass $1/K$ has been added at location x_1 (for ℓ large enough). If $J \geq 2$, then we again bring mass from $x_{\ell+1}, \ldots$, until another mass $1/K$ has been added at x_2. We carry on until all the mass at x_J has been used, thus building a transport plan $(\pi_{ij})_{1\leq i,j\leq N}$ which sends $\tilde{\mu}$ onto μ', in such a way that $\pi_{ii} \geq \dfrac{n_i}{K}$ for all i. Hence,

$$\sum_{j\neq i} \pi_{ij} \leq \beta_i - \pi_{ii} = \beta_i - \frac{n_i}{K} \leq \frac{1}{K},$$

and this plan yields an upper bound on the Wasserstein distance :

$$W_p^p(\tilde{\mu}, \mu') \leq \sum_{i,j=1}^N d(x_i, x_j)^p\, \pi_{ij} = \sum_{i=1}^N \sum_{j\neq i} d(x_i, x_j)^p\, \pi_{ij} \leq N \frac{D^p}{K}.$$

To summarize the first two steps : for any μ in $\mathcal{P}(E)$ there exists $\mu' \in \mathcal{C}_K$ such that

$$W_p(\mu, \mu') \leq\; r + D \left(\frac{N}{K}\right)^{1/p}.$$

In other words, the family $\Big(B\big(\mu', r + D(N/K)^{1/p}\big)\Big)_{\mu'\in\mathcal{C}_K}$ covers $\mathcal{P}(E)$.

In the **third step** we choose some suitable K and r for a given δ.

We first choose K in such a way that r and $D(N/K)^{1/p}$ have the same order of magnitude, for instance

$$K = \left[N \left(\frac{D}{r}\right)^p \right] + 1.$$

Then

$$r + D\,(N/K)^{1/p} \leq 2r,$$

and the balls $B\big(\mu', r + D(N/K)^{1/p}\big)$ have radius at most δ if

$$r = \frac{\delta}{2}.$$

Now K and r are fixed, $N = N(E, \delta/2)$, and we just have to estimate the cardinality $\sharp\mathcal{C}_K$ of $\mathcal{C}_K$. For this we first note that

$$\sharp\mathcal{C}_K = \frac{(K+N-1)!}{(K-1)!N!} = \frac{(K+N-1)\dots K}{N!}$$

Without loss of generality, we have assumed $\delta < D$, so $K > N$. Then $K < \cdots < K+N-1 < 2K$, and hence

$$\sharp\mathcal{C}_K \leq \frac{(2K)^N}{N!} \leq \left(\frac{2Ke}{N}\right)^N.$$

Since $N \geq 1$ and $2D \geq \delta$, we can write

$$K \leq N\left(\frac{2D}{\delta}\right)^p + 1 \leq 2N\left(\frac{2D}{\delta}\right)^p,$$

and we deduce

$$\sharp\mathcal{C}_K \leq \left(C\,\frac{D}{\delta}\right)^{pN\left(E,\frac{\delta}{2}\right)}$$

with $C = 2(4e)^{1/p} \leq 8e$.

Consequently, we have covered $\mathcal{P}(E)$ by the $\left(C\frac{D}{\delta}\right)^{pN\left(E,\frac{\delta}{2}\right)}$ balls $(B(\mu', \delta))_{\mu' \in \mathcal{C}_K}$ with radius δ. This concludes the argument. □

In the particular case when E is the Euclidean ball B_R of radius R in $\mathbb{R}^d$, we have

$$N(B_R, r) \leq k\left(\frac{R}{r}\right)^d \tag{6.86}$$

for some constant k. To see this, one may for instance consider the balls with center in the lattice $\frac{r}{\sqrt{d}}\mathbb{Z}^d$ in $\mathbb{R}^d$. Then Theorem 6.25 yields the bound

$$\mathcal{N}_p(B_R, \delta) \leq \left(C\,\frac{R}{\delta}\right)^{pk\left(\frac{R}{\delta}\right)^d},$$

which is used in the present paper.

6.8 Appendix : regularity estimates on the limit PDE

In this appendix we study solutions to the limit equation

$$\frac{\partial \mu}{\partial t} = \Delta\mu + \nabla\cdot(\mu\nabla(V + W * \mu)), \qquad t \geq 0, \quad x \in \mathbb{R}^d \tag{6.87}$$

and establish the regularity results stated in Proposition 6.19. Following the method in [44], we shall measure the regularity in terms of L^2-Sobolev spaces

$$H^s(\mathbb{R}^d) = \Big\{u \in L^2(\mathbb{R}^d); \partial^\alpha u \in L^2(\mathbb{R}^d),\, \alpha \in \mathbb{N}^d,\, |\alpha| \le s\Big\} \qquad (s \in \mathbb{N}).$$

Our main result is as follows.

Theorem 6.27. *Let V and W such that all their partial derivatives $\partial^\alpha V$ and $\partial^\alpha W$ are continuous and grow at most polynomially at infinity, for any multi-index $\alpha \in \mathbb{N}^d$ with $|\alpha| \le s+1$. Let $a, E > 0$ and let μ_0 be a probability density such that*

$$\int_{\mathbb{R}^d} e^{a|x|^2}\, d\mu_0(x) \le E.$$

Then, there exists a continuous function $f : (0,+\infty) \to (0,+\infty)$, only depending on d, s, V, W, a and E, such that any classical solution $\mu = \mu(t,x)$ to (6.87), starting from μ_0, satisfies

$$\big\|\mu(t,\cdot)\big\|_{H^s(\mathbb{R}^d)} \le f(t).$$

Proof. For the sake of simplicity we only give a formal proof, which can be turned rigorous by means of regularization arguments.

Let then $\mu = (\mu(t,.))_{t\ge 0}$ be a solution of

$$\frac{\partial \mu}{\partial t} = \Delta\mu + \nabla\cdot(\mu\nabla(V + W*\mu)), \qquad t \ge 0, \quad x \in \mathbb{R}^d;$$

we rewrite the equation as

$$\partial_t \mu = \sum_{i=1}^d \partial_{ii}\mu + \partial_i\left[\mu\partial_i\phi\right],$$

where $\partial_i = \partial^{e_i}$ if e_i is the i-th vector of the canonical base of $\mathbb{R}^d$, and

$$\phi(t,x) = V(x) + W*\mu(t,x).$$

Let $\alpha \in \mathbb{N}^d$ be given. By integration by parts and Cauchy-Schwarz inequality,

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^d}|\partial^\alpha\mu|^2 &= \int_{\mathbb{R}^d}\partial^\alpha\mu\,\partial_t(\partial^\alpha\mu) = \int_{\mathbb{R}^d}\partial^\alpha\mu\,\partial^\alpha(\partial_t\mu)\\
&= \sum_{i=1}^d\int_{\mathbb{R}^d}\partial^\alpha\mu\,\partial^\alpha\left(\partial_{ii}\mu + \partial_i\left[\mu\,\partial_i\phi\right]\right)\\
&= -\sum_{i=1}^d\int_{\mathbb{R}^d}\left|\partial^{\alpha+e_i}\mu\right|^2 + \int\partial^{\alpha+e_i}\mu\,\partial^\alpha\left[\mu\partial_i\phi\right]\\
&\le -\sum_{i=1}^d\int_{\mathbb{R}^d}\left|\partial^{\alpha+e_i}\mu\right|^2 + \sum_i^d\left[\int_{\mathbb{R}^d}\left|\partial^{\alpha+e_i}\mu\right|^2\right]^{1/2}\left[\sum_{\beta\le\alpha}C_{\alpha,\beta}\int_{\mathbb{R}^d}\left|\partial^{\alpha-\beta+e_i}\phi\,\partial^\beta\mu\right|^2\right]^{1/2}\\
&\le -\frac{1}{2}\sum_{i=1}^d\int_{\mathbb{R}^d}\left|\partial^{\alpha+e_i}\mu\right|^2 + \sum_{\beta\le\alpha}C_{\alpha,\beta}\int_{\mathbb{R}^d}\left|\partial^{\alpha-\beta+e_i}\phi\,\partial^\beta\mu\right|^2.
\end{aligned}$$

By summing over $\alpha \in \mathbb{N}^d$ with $|\alpha| \le s$, we find

$$\frac{d}{dt}\sum_{|\alpha|\le s}\int_{\mathbb{R}^d}|\partial^\alpha\mu|^2 \le -\sum_{|\alpha|\le s}\sum_{i=1}^d\int_{\mathbb{R}^d}\left|\partial^{\alpha+e_i}\mu\right|^2 + \sum_{|\alpha|\le s}\sum_{\beta\le\alpha}C_{\alpha,\beta}\int_{\mathbb{R}^d}\left|\partial^{\alpha-\beta+e_i}\phi\,\partial^\beta\mu\right|^2.$$

Given $T>0$, by Proposition 6.14 there exist constants $\hat{a}$ and $\hat{E}$, depending only on d, a, E and T, such that

$$\int e^{\hat{a}|x|^2}d\mu(t,x) \le \hat{E} \tag{6.88}$$

for all $t\in[0,T]$. In particular, it follows from our assumptions on the derivatives of V and W that all $\left|\partial^{\alpha-\beta+e_i}\phi\right|^2$ terms are bounded by some polynomial in $|x|$, uniformly in $t\in[0,T]$.

Let $\langle x\rangle := \sqrt{1+|x|^2}$. For $k,s\ge 0$, we introduce the weighted norms

$$\|u\|_{H^s_k} := \left(\sum_{|\alpha|\le s}\int_{\mathbb{R}^d}\langle x\rangle^k\,|\partial^\alpha u(x)|^2\,dx\right)^{1/2}$$

and

$$\|u\|_{L^1_k} := \int_{\mathbb{R}^d}\langle x\rangle^k\,|u(x)|\,dx.$$

Then for any $s\in\mathbb{N}$ and $T\ge 0$ there exist k and $C\ge 0$ such that

$$0\le t\le T \Longrightarrow \qquad \frac{d}{dt}\|\mu\|^2_{H^s} \le -\|\mu\|^2_{H^{s+1}} + C\,\|\mu\|^2_{H^s_k}. \tag{6.89}$$

We shall prove later on the following interpolation lemma :

Lemma 6.28. *Given $d\ge 1$, $s\in\mathbb{N}$ an $k\ge 0$, there exist nonnegative constants $C(d,s,k)$ and $h(d,s,k)$, and $\theta(d,s)\in(0,1)$ such that for all $u\in L^1_\infty(\mathbb{R}^d)\cap H^{s+1}(\mathbb{R}^d)$,*

$$\|u\|_{H^s_k} \le C(d,s,k)\|u\|^{1-\theta(d,s)}_{L^1_{h(d,s,k)}}\,\|u\|^{\theta(d,s)}_{H^{s+1}}.$$

Then, again from (6.88), all $\|\mu\|_{L^1_{h(d,s,k)}}(t)$ norms are bounded on $[0,T]$, so from (6.89) and Lemma 6.28 there exists some constants C such that

$$\frac{d}{dt}\|\mu\|^2_{H^s} \le -\|\mu\|^2_{H^{s+1}} + C\,\|\mu\|^{2\theta}_{H^{s+1}} \le -\frac{1}{2}\|\mu\|^2_{H^{s+1}} + C \le -C\|\mu\|^{2/\theta}_{H^s} + C.$$

In other words $A(t)=\|\mu\|^2_{H^s}(t)$ satisfies on $[0,T]$ the differential inequality

$$A'(t) + c\,A(t)^p \le C \tag{6.90}$$

for some constants $c,C\ge 0$ and $p=1/\theta>1$ depending only on d, a, E, s and T.

Let us distinguish two cases. If $A(0) \leq 1$, then we only use the inequality $A'(t) \leq C$ to make sure that

$$A(t) \leq A(0) + Ct \leq 1 + CT$$

for any $t \in [0, T]$.

If on the other hand $A(0) \geq 1$, we deduce from (6.90) that

$$A'(t) + c\,A(t)^p \leq CA(t),$$

as long as $A(t) \geq 1$, so that $D(t) := A(t)^{1-p}$ satisfies the inequality

$$D'(t) + (p-1)\,CD(t) \geq (p-1)\,c$$

which integrates to

$$D(t) \geq D(0)e^{(1-p)Ct} + \frac{c}{C}(1 - e^{(1-p)Ct}) \geq \frac{c}{C}(1 - e^{(1-p)t}).$$

As a consequence, as long as $A(t) \geq 1$, we have

$$A(t) \leq (c/C)^{1/1-p}(1 - e^{(1-p)t})^{1/(1-p)}.$$

In the end, we have obtained an a priori bound on $A(t) = \displaystyle\int |\partial^\alpha \mu|^2\,(t)$ for $t \in (0, T]$, depending only on d, s, a, E and T, but not on the initial value $A(0)$. Then the proof can be concluded by an approximation argument. □

Proof of Lemma 6.28. We proceed by induction on s.

In the **first step** we prove the result for $s = 0$. Given $d \geq 1$ and $a \in (0, 1]$, we write

$$\int_{\mathbb{R}^d} \langle x\rangle^k\,|u(x)|^2\,dx = \int_{\mathbb{R}^d} \langle x\rangle^k\,|u(x)|^a\,|u(x)|^{2-a}\,dx,$$

so, by Hölder's inequality,

$$\|u\|^2_{L^2_k} \leq \|u\|^a_{L^1_{\frac{k}{a}}}\;\|u\|^{2-a}_{L^{\frac{2-a}{1-a}}}$$

(with $\dfrac{2-a}{1-a} = \infty$ if $a = 1$). Then by Sobolev embedding,

$$\|u\|^2_{L^2_k} \leq C(d, a)\|u\|^a_{L^1_{\frac{k}{a}}}\;\|u\|^{2-a}_{H^1},$$

where $a = 1$ if $d = 1$, a is arbitrary in $(0, 1)$ if $d = 2$, and $a = \dfrac{4}{d+2}$ if $d \geq 3$, that is,

$$\|u\|_{L^2_k} \leq C(d)\|u\|^{1-\theta(d)}_{L^1_{\frac{k}{a}}}\;\|u\|^{\theta(d)}_{H^1}$$

where $\theta(1) = \dfrac{1}{2}$, any $\theta(2) \in \left(\dfrac{1}{2}, 1\right)$ for $d = 2$, and $\theta(d) = \dfrac{d}{d+2}$ for $d \geq 3$.

In the **second step** we let $s \geq 1$ and assume by induction that there exist some constants $C(d,s-1,k), h(d,s-1,k) \geq 0$ and $\theta(d,s-1) \in (0,1)$ such that for all $u \in L^1_\infty(\mathbb{R}^d) \cap H^s(\mathbb{R}^d)$:

$$\|u\|_{H^{s-1}_k} \leq C(d,s-1,k) \|u\|^{1-\theta(d,s-1)}_{L^1_{h(d,s-1,k)}} \|u\|^{\theta(d,s-1)}_{H^s}.$$

Let then $u \in L^1_\infty(\mathbb{R}^d) \cap H^{s+1}(\mathbb{R}^d)$.

Given $\alpha \in \mathbb{N}^d$ with $|\alpha| = j$ and $1 \leq j \leq s$, we split α into $\alpha = \alpha_1 + \alpha_2$ with $|\alpha_2| = 1$, and integrate by parts :

$$\begin{aligned} \|\partial^\alpha u\|^2_{L^2_k} &\leq k \, \|\partial^{\alpha_1} u\|_{L^2_{2k-2}} \, \|\partial^\alpha u\|_{L^2} + \|\partial^{\alpha_1} u\|_{L^2_{2k}} \, \|\partial^{\alpha+\alpha_2} u\|_{L^2} \\ &\leq (k+1) \|\partial^{\alpha_1} u\|_{L^2_{2k}} \sup_{|\alpha| \leq j+1} \|\partial^\alpha u\|_{L^2}, \end{aligned}$$

whence

$$\begin{aligned} \sup_{|\alpha|=j} \|\partial^\alpha u\|^2_{L^2_k} &\leq (k+1) \sup_{|\alpha|=j-1} \|\partial^\alpha u\|_{L^2_{2k}} \sup_{|\alpha|\leq j+1} \|\partial^\alpha u\|_{L^2} \\ &\leq (k+1) \sup_{|\alpha|\leq s-1} \|\partial^\alpha u\|_{L^2_{2k}} \sup_{|\alpha|\leq s+1} \|\partial^\alpha u\|_{L^2}. \end{aligned}$$

Since this holds for any $1 \leq j \leq s$ we obtain

$$\sup_{1\leq|\alpha|\leq s} \|\partial^\alpha u\|^2_{L^2_k} \leq (k+1) \sup_{|\alpha|\leq s-1} \|\partial^\alpha u\|_{L^2_{2k}} \sup_{|\alpha|\leq s+1} \|\partial^\alpha u\|_{L^2}.$$

Moreover

$$\|u\|^2_{L^2_k} \leq \|u\|_{L^2_{2\,k}} \, \|u\|_{L^2} \leq \sup_{|\alpha|\leq s-1} \|\partial^\alpha u\|_{L^2_{2k}} \sup_{|\alpha|\leq s+1} \|\partial^\alpha u\|_{L^2},$$

so that finally

$$\|u\|^2_{H^s_k} \leq (k+1) \|u\|_{H^{s-1}_{2k}} \, \|u\|_{H^{s+1}}.$$

Then, by induction hypothesis,

$$\|u\|^2_{H^s_k} \leq (k+1)\, C(d,s-1,2k) \|u\|^{1-\theta(d,s-1)}_{L^1_{h(d,s-1,2k)}} \, \|u\|^{\theta(d,s-1)}_{H^s} \, \|u\|_{H^{s+1}},$$

whence

$$\|u\|_{H^s_k} \leq C(d,k,s) \|u\|^{1-\theta(d,s)}_{L^1_{h(d,s,k)}} \, \|u\|^{\theta(d,s)}_{H^{s+1}}$$

where $\theta(d,s) = \dfrac{1}{2-\theta(d,s-1)} \in (0,1)$ and $h(d,s,k) = h(d,s-1,2k) \geq 0$. This concludes the argument. $\square$

Chapitre 7

Concentration de la mesure empirique sur les trajectoires des particules

Ce chapitre correspond à l'article [20].

Dans ce chapitre nous prolongeons l'étude de la limite de champ moyen d'un système de particules abordée dans le chapitre pécédent. Nous obtenons en particulier des estimations de grandes déviations non asymptotiques mesurant la concentration autour de sa limite de la mesure empirique des particules, non plus à un instant donné comme précédemment, mais au niveau des trajectoires. Pour cela nous adaptons à des espaces de trajectoires les techniques développées dans le chapitre précédent pour $\mathbb{R}^d$.

Introduction

This paper is devoted to the study of the behaviour of some large stochastic particle system. In the models to be considered, the evolution of each particle is governed by a random diffusive term, an exterior force field and a mean field interaction with the other particles. For such systems the limit behaviour has been clearly identified and studied in terms of law of large numbers, central limit theorem and large deviations. Here we shall give new quantitative estimates on the convergence in question in the setting of large deviations.

This follows some works addressing this issue at the level of observables or at the level of the whole system at a given time, that we now summarize. For that purpose let $(X_t^i)_{1\leq i\leq N}$ be the position at time t of the N particles of the system in the phase space $\mathbb{R}^d$ and let μ_t be some probability measure describing the limit behaviour of the system. At the level of Lipschitz observables, F. Malrieu [74] adapted ideas of concentration of measure to obtain bounds like

$$\sup_{[\varphi]_1\leq 1} \mathbb{P}\left[\Big|\frac{1}{N}\sum_{i=1}^N \varphi(X_t^i) - \int_{\mathbb{R}^d} \varphi\, d\mu_t\Big| > \frac{C}{\sqrt{N}} + \varepsilon\right] \leq 2\, e^{-\lambda N \varepsilon^2}, \qquad N \geq 1, \tag{7.1}$$

where C and λ are some constants independent of ε and N, and $[\,\cdot\,]_1$ is the Lipschitz seminorm defined by

$$[\varphi]_1 := \sup_{x\neq y} \frac{|\varphi(x)-\varphi(y)|}{|x-y|}.$$

In other words, letting δ_x stand for the Dirac mass at a point $x \in \mathbb{R}^d$, the empirical measure

$$\hat{\mu}_t^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$$

of the system, which generates the observables at time t, satisfies the deviation inequality

$$\sup_{[\varphi]_1\leq 1} \mathbb{P}\left[\Big|\int_{\mathbb{R}^d} \varphi\, d\hat{\mu}_t^N - \int_{\mathbb{R}^d} \varphi\, d\mu_t\Big| > \frac{C}{\sqrt{N}} + \varepsilon\right] \leq 2\, e^{-\lambda N \varepsilon^2}, \qquad N \geq 1.$$

Now one can measure how this empirical measure $\hat{\mu}_t^N$ is close to its limit μ_t in a stronger sense, namely, at the very level of the measures. For this, adapting Sanov's large deviation argument, the authors in [22] got quantitative and non-asymptotic bounds on the deviation of $\hat{\mu}_t^N$ around μ_t for some distance which induces a topology stronger that the narrow topology. By comparison with (7.1), these bounds can be written as

$$\mathbb{P}\left[\sup_{[\varphi]_1\leq 1}\Big|\frac{1}{N}\sum_{i=1}^N \varphi(X_t^i) - \int_{\mathbb{R}^d} \varphi\, d\mu_t\Big| > \varepsilon\right] \leq C(\varepsilon)\, e^{-\lambda N \varepsilon^2}, \qquad N \geq 1. \tag{7.2}$$

In this work we want to go one step further by considering the whole **trajectories** of the particles. A natural object to consider is the empirical measure of the trajectories $(X_t^i)_{0\leq t\leq T}$ on some given time interval $[0, T]$, which is defined as

$$\hat{\mu}_{[0,T]}^N := \frac{1}{N}\sum_{i=1}^N \delta_{(X_t^i)_{0\leq t\leq T}}$$

where $\delta_{(X_t^i)_{0\leq t\leq T}}$ is the Dirac mass on the path $(X_t^i)_{0\leq t\leq T}$. This is a random probability measure, no longer on the phase space $\mathbb{R}^d$, but now on the path space, which in our model is the space of $\mathbb{R}^d$-valued continuous functions on $[0, T]$.

Its limit behaviour is given as follows : the limit μ_t of the empirical measure $\hat{\mu}_t^N$ at time t can be seen as the law of the solution at time t to a (nonlinear) stochastic differential equation; then the law of the whole process on $[0, T]$ so defined will be the limit of $\hat{\mu}_{[0,T]}^N$. We shall give a precise meaning to this convergence, and obtain some estimates which are the analogue of (7.2) in the path space; in particular we shall see that they imply (7.2) by projection at time t.

In the next section we state our main results and give an insight of the proofs, which will be given in more detail in the following sections.

7.1 Statement of the results

7.1.1 Some notation and definitions

One of the key points in this work is to measure the discrepancy between probability measures : this will be done by means of **transportation** (or Wasserstein) **distances**, which have revealed convenient in this type of issues and are defined as follows. Let (X, d) be a separable and complete metric space, and let p be a real number, $p \geq 1$; then the Wasserstein distance of order p between two Borel probability measures μ and ν on X is

$$W_p(\mu, \nu) := \inf_{\pi} \left(\iint_{X \times X} d(x, y)^p \, d\pi(x, y) \right)^{1/p}$$

where π runs over the set of all joint measures on $X \times X$ with marginals μ and ν. It can be checked that W_p induces a metric on the set of Borel probability measures on X such that the moment $\displaystyle\int_X d(x_0, x)^p \, d\mu(x)$ be finite for some (and thus any) x_0 in X ; convergence in this metric is equivalent to narrow convergence (against bounded continuous functions) plus some tightness condition on the moments (see for instance [111] for further details on these distances).

At some point the space (X, d) will be $\mathbb{R}^d$ equipped with the Euclidean distance $|\cdot|$, and in this context W_p shall be denoted $W_{p,\tau}$. But (X, d) will mainly be the space $\mathcal{C}([0, T], \mathbb{R}^d)$, also denoted $\mathcal{C}$ if no confusion is possible, of $\mathbb{R}^d$-valued continuous functions on $[0, T]$, equipped with the uniform norm

$$\|f\|_\infty := \sup_{0 \leq t \leq T} |f(t)|;$$

for this space the Wasserstein distances will be denoted $W_{p,[0,T]}$. The Wasserstein distances considered in these two situations are linked in the following way : if, for $0 \leq t \leq T$, π_t is the projection from $\mathcal{C}$ into $\mathbb{R}^d$ defined by $\pi_t(f) = f(t)$, then for any Borel probability measures μ and ν on $\mathcal{C}$, and any $p \geq 1$, the relation

$$W_{p,\tau}(\pi_t \sharp \mu, \pi_t \sharp \nu) \leq W_{p,[0,T]}(\mu, \nu), \qquad 0 \leq t \leq T \tag{7.3}$$

holds, where $\pi_t \sharp \mu$ is the image measure of μ by π_t.

The distance between two probability measures μ and ν on X can also be expressed in terms of the **relative entropy** of ν with respect to μ (for instance), defined by

$$H(\nu|\mu) = \int_X \frac{d\nu}{d\mu} \ln \frac{d\nu}{d\mu} \, d\mu$$

if ν is absolutely continuous with respect to μ, and $H(\nu|\mu) = +\infty$ otherwise.

Both notions are linked by the family of *transportation* or *Talagrand inequalities* : given $p \geq 1$ and $\lambda > 0$, we say that a probability measure μ on X satisfies the inequality $T_p(\lambda)$ if

$$W_p(\mu, \nu) \leq \sqrt{\frac{2}{\lambda} H(\nu|\mu)}$$

holds true for any measure ν, and we say that μ satisfies T_p if it satisfies $T_p(\lambda)$ for some $\lambda > 0$. By Jensen's inequality, the weakest of all is T_1, which is also the only one for which a simple characterization is known : a measure μ satisfies $T_1(\lambda)$ for some $\lambda > 0$ if and only if it admits a square exponential moment, in the sense that there exist some $a > 0$ and x_0 in X such that $\int_X \exp\left(a\, d(x_0, x)^2\right) d\mu(x)$ be finite. Numerical relations between such a and λ can be found in [23, 46].

7.1.2 A general concentration inequality for empirical measures

The proof of our main theorem on the particle system is based on some general concentration result for the empirical measure of $\mathcal{C}$-valued independent and identically distributed random variables. We state this result separately.

For this purpose, given some Borel probability measure μ on $\mathcal{C}$, and N independent random variables $(X^i)_{1 \le i \le N}$ with law μ, we denote

$$\hat{\mu}^N := \frac{1}{N} \sum_{i=1}^{N} \delta_{X^i}$$

their empirical measure.

Given some real number $\alpha \in (0, 1]$, we denote $\mathcal{C}^\alpha := \mathcal{C}^\alpha([0, T], \mathbb{R}^d)$ the space of functions in $\mathcal{C} := \mathcal{C}([0, T], \mathbb{R}^d)$ which moreover are Hölder of order α, equipped with the Hölder norm

$$\|f\|_\alpha := \sup\left(\|f\|_\infty, [f]_\alpha\right)$$

where

$$[f]_\alpha := \sup_{0 \le t, s \le T} \frac{|f(t) - f(s)|}{|t - s|^\alpha}.$$

$\mathcal{C}^\alpha$ is a Borel set of the space $\mathcal{C}$ equipped with the topology induced by the uniform norm, and for Borel measures on $\mathcal{C}$, concentrated on $\mathcal{C}^\alpha$, we have in the above notation :

Theorem 7.1. *Let $p \in [1, 2]$ and let μ be a Borel probability measure on $\mathcal{C}$ satisfying a $T_p(\lambda)$ inequality for some $\lambda > 0$, and such that $\int_{\mathcal{C}} e^{a\|x\|_\alpha^2}\, d\mu(x)$ be finite for some $a > 0$ and $\alpha \in (0, 1]$. Then, for any $\alpha' < \alpha$ and $\lambda' < \lambda$, there exists some constant N_0 such that*

$$\mathbb{P}\left[W_{p,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \le e^{-\beta_p \frac{\lambda'}{2} N \varepsilon^2}, \tag{7.4}$$

for any $\varepsilon > 0$ and $N \ge N_0\, \varepsilon^{-2} \exp\left(N_0\, \varepsilon^{-1/\alpha'}\right)$, where

$$\beta_p = \begin{cases} 1 & \text{if } \; 1 \le p < 2 \\ (1 + \sqrt{\lambda/a})^{-2} & \text{if } \quad p = 2. \end{cases}$$

Here the constants N_0 depends on μ only through λ, a, α and $\displaystyle\int_{\mathcal{C}} e^{a\|x\|_\alpha^2}\, d\mu(x)$.

Let us make a few remarks on this result.

First of all, for another formulation of the obtained bound, we recall Kantorovich-Rubinstein dual expression of the W_1 distance on a general space (X, d) :

$$W_1(\mu,\nu) = \sup_{[\varphi]_1 \le 1} \Big\{ \int_X \varphi\, d\mu - \int_X \varphi\, d\nu \Big\} \tag{7.5}$$

where $[\varphi]_1 := \sup_{x\neq y} \dfrac{|\varphi(x)-\varphi(y)|}{d(x,y)}\cdot$ Then a result by S. Bobkov and F. Götze [17] ensures that a $T_1(\lambda)$ inequality for μ is equivalent to the concentration inequality

$$\sup_{[\varphi]_1\le 1} \mathbb{P}\left[\frac{1}{N}\sum_{i=1}^N \varphi(X^i) - \int_X \varphi\, d\mu > \varepsilon \right] \le e^{-\frac{\lambda}{2}N\varepsilon^2}, \qquad N \ge 1\,.$$

By comparison, the bound given by Theorem 7.1 implies, by (7.5),

$$\mathbb{P}\left[\sup_{[\varphi]_1\le 1} \Big(\frac{1}{N}\sum_{i=1}^N \varphi(X^i) - \int_{\mathcal{C}} \varphi\, d\mu \Big) > \varepsilon \right] \le e^{-\frac{\lambda'}{2}N\varepsilon^2}, \quad \lambda' < \lambda\,, \quad N \text{ large enough}\,,$$

but a modification of the proof would also lead to

$$\mathbb{P}\left[\sup_{[\varphi]_1\le 1} \Big(\frac{1}{N}\sum_{i=1}^N \varphi(X^i) - \int_{\mathcal{C}} \varphi\, d\mu \Big) > \varepsilon \right] \le C(\varepsilon)\, e^{-\frac{\lambda'}{2}N\varepsilon^2}, \quad \lambda' < \lambda\,, \quad N \ge 1, \tag{7.6}$$

for some computable large constant $C(\varepsilon)$. In other words we control a much stronger quantity, up to some loss on the constant in the right-hand side, or some condition on the size of the sample.

This result seems reasonable in view of Sanov's theorem (stated in [43] for instance). By applying this theorem to $A := \{\nu; W_{p,[0,T]}(\nu,\mu) \ge \varepsilon\}$, for some given $\varepsilon > 0$, one can hope for a bound like

$$\mathbb{P}\left[W_{p,[0,T]}(\mu, \hat\mu^N) \ge \varepsilon \right] \le \exp\Big(-N \inf\big\{ H(\nu|\mu); \nu \in A \big\} \Big)$$

for large N. With this bound in hand, since

$$\inf\big\{ H(\nu|\mu); \nu\in A\big\} \ge \frac{\lambda}{2}\,\varepsilon^2$$

as μ satisfies $T_p(\lambda)$, one indeed obtains an upper bound like (7.4), but only in an asymptotic way, whereas Theorem 7.1 moreover gives an estimate on a sufficient size of the sample for the deviation bound to hold. Sanov's theorem does not actually give such a relevant upper bound here ; indeed, on an unbounded space such as $\mathcal{C}$, the closure $\overline{A}$ of A (for the narrow topology) contains μ itself : in particular $\inf\{H(\nu|\mu);\ \nu\in\overline{A}\} = 0$ and Sanov's theorem only gives the trivial upper bound $\mathbb{P}\,[A] \le \exp\big(-N \inf\{H(\nu|\mu); \nu\in\overline{A}\}\big) = 1$.

To get a more relevant upper bound we impose some extra integrability assumption that may at first sight seem strong and odd. The reason is that, proceeding as in [22], we first reduce the problem to a compact set of $\mathcal{C}$ that has almost full μ measure : a large ball of $\mathcal{C}^\alpha$ will do by Ascoli's theorem and the integrability assumption on μ. Then on this compact set one can get precise upper bounds by using some techniques based on some covering argument and developed in [66] (see also [43, Exercises 4.4.5 and 6.2.19] and [56]). And actually the assumption is satisfied by the Wiener measure on $\mathcal{C}$ (recall that the Brownian motion paths are almost surely Hölder of order α for any $\alpha < 1/2$) and by extension by the law of the process we shall be considering.

This integrability assumption again implies the existence of a square-exponential moment for μ on $\mathcal{C}$ (for the uniform norm). Since this is equivalent to some T_1 inequality for μ, the $T_p(\lambda)$ assumption is redondant in the case when $p = 1$ if one does not care of the involved constants.

Finally this result can be seen as an extension of the following similar concentration result given in [22, Theorem 1.1] in the case of measures on $\mathbb{R}^d$: if m satisfies some $T_p(\lambda)$, then

$$\mathbb{P}\left[W_{p,\tau}(m, \hat{m}^N) > \varepsilon\right] \leq e^{-\gamma_p \frac{\lambda'}{2} N \varepsilon^2}, \quad \varepsilon > 0\,, \quad N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1). \tag{7.7}$$

Let indeed m be such a measure on $\mathbb{R}^d$. Then the law μ of a constant process on $[0, T]$ initially distributed according to m satisfies the assumptions of Theorem 7.1 (one can take any $a < \lambda/2$), and the bound (7.7) follows by (7.3) with the constant γ_p obtained in [22]. Note however that the required size of the sample is here much larger for small ε.

This theorem will be proved in Section 7.2.

7.1.3 Interacting particle systems

We now turn to the study of a system of N stochastic interacting particles which positions X_t^i in the phase space $\mathbb{R}^d$ $(1 \leq i \leq N)$ evolve according to the system of coupled stochastic differential equations

$$dX_t^i = \sqrt{2}\, dB_t^i - \nabla V(X_t^i)dt - \frac{1}{N}\sum_{j=1}^N \nabla W(X_t^i - X_t^j)dt, \qquad 1 \leq i \leq N. \tag{7.8}$$

Here the B^i's are N standard independent Brownian motions on $\mathbb{R}^d$, V and W are exterior and interaction potentials respectively.

The state of the system at some given time t can be described by some random probability measure on the phase space $\mathbb{R}^d$, defined as

$$\hat{\mu}_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$$

and called the empirical measure of the system. As pointed out in the introduction, the observables of the system are simply given by

$$\frac{1}{N}\sum_{i=1}^N \varphi(X_t^i) = \int_{\mathbb{R}^d} \varphi(x)\, d\hat{\mu}_t^N(x),$$

whence the relevance of this object. Another remarkable feature of this measure is that it belongs to the same functional space (of measures on $\mathbb{R}^d$), independently of the size of the system. Hence there is some hope for the empirical measure to converge when N goes to infinity. And indeed, under some assumptions on the potential V and W, it has been proven that if the particles are initially distributed in a chaotic way, for instance as independent and identically distributed variables, then $\hat{\mu}_t^N$ converges as N tends to $+\infty$ to a solution to the partial differential equation

$$\frac{\partial \mu_t}{\partial t} = \Delta \mu_t + \nabla \cdot \big(\mu_t \nabla (V + W * \mu_t)\big). \tag{7.9}$$

This nonlinear diffusive equation, in which $\Delta, \nabla\cdot$ and ∇ respectively stand for the Laplace, divergence and gradient operators, is an instance of a McKean-Vlasov equation and has been used in the modelling of one-dimensional granular media in [11] for instance. The convergence of the empirical measure is strongly linked with the phenomenon of propagation of chaos for the interacting particles, and both issues have been studied by H. Tanaka [109], A.-S. Sznitman [105], S. Méléard [84] or S. Benachour, B. Roynette, D. Talay and P. Vallois [9, 10] for instance, under various assumptions on the potentials V and W and in different senses. Then quantitative concentration estimates, in the setting of large deviations, of the empirical measure around its limit μ_t have been obtained by F. Malrieu [74] at the level of observables, and later at the level of the law itself in [22].

In this work we want to go one step further by studying the limit behaviour of the empirical measure

$$\hat{\mu}_{[0,T]}^N = \frac{1}{N} \sum_{i=1}^N \delta_{X^i}$$

of the trajectories $X^i = (X_t^i)_{0 \leq t \leq T}$ of the particles on some time interval $[0, T]$.

For that purpose, let $Y = (Y_t)_{0 \leq t \leq T}$ be a solution to the equation

$$dY_t = \sqrt{2}\, dB_t - \nabla V(Y_t)\, dt - \nabla W * \nu_t(Y_t)\, dt \tag{7.10}$$

starting at Y_0 distributed according to the initial condition μ_0 in (7.9), where ν_t is the law of Y_t at time t. Then, by Itô's formula, ν_t also is a solution to equation (7.9) with initial datum μ_0, and a uniqueness result ensures that actually $\nu_t = \mu_t$. In other words the limit behaviour μ_t of $\hat{\mu}_t^N$ is the time-marginal of the law $\mu_{[0,T]}$ of the process Y, in the sense that it is the image measure of $\mu_{[0,T]}$ by the canonical projection π_t defined on the paths space $\mathcal{C}$ by $\pi_t(f) = f(t)$. Since $\hat{\mu}_t^N$ is the time-marginal of the empirical measure $\hat{\mu}_{[0,T]}^N$ that we are considering, it is natural to hope that $\hat{\mu}_{[0,T]}^N$ converge (in some sense) to $\mu_{[0,T]}$. This convergence has indeed been proved in the first works mentionned above, and here we want to extend to this new framework the techniques that have been developed in [74] and more particularly in [22].

We shall assume that the potentials V and W are twice differentiable on $\mathbb{R}^d$, with bounded hessian matrices in the sense that there exist some real constants β, β', γ and γ' such that

$$\beta\, I \leq D^2 V(x) \leq \beta'\, I, \qquad \gamma\, I \leq D^2 W(x) \leq \gamma'\, I, \qquad x \in \mathbb{R}^d. \tag{7.11}$$

In other words the force fields ∇V and ∇W are assumed to be Lipschitz on the whole $\mathbb{R}^d$.

Under these assumptions, global existence and uniqueness, pathwise and in law, of the solutions to (7.8) and (7.10) are proven in [84] for instance for square-integrable initial data; moreover the paths are continuous (in time). We shall also assume that the potential W, which gives rise to an interaction term, is symmetric in the sense that $W(-z) = W(z)$ for all $z \in \mathbb{R}^d$. Then we shall prove

Theorem 7.2. *Let μ_0 be a probability measure on $\mathbb{R}^d$, admitting a finite square-exponential moment in the sense that there exists some $a_0 > 0$ such that $\int_{\mathbb{R}^d} e^{a_0|x|^2}\, d\mu_0(x)$ be finite. Let $(X_0^i)_{1\le i\le N}$ be N independent random variables with common law μ_0. Given $T \ge 0$, let $(X^i)_i$ be the solution of (7.8) on $[0,T]$ with initial value $(X_0^i)_i$, where V and W are assumed to satisfy (7.11); let also $\hat\mu^N_{[0,T]}$ be the empirical measure associated with the N paths X^i. Let finally $\mu_{[0,T]}$ be the law of the process solution of (7.10) for some initial value distributed according to μ_0.*

Then, for any $\alpha \in (0,1/2)$, there exist some positive constants K and N_0 such that

$$\mathbb{P}\,\left[W_{1,[0,T]}(\mu_{[0,T]}, \hat\mu^N_{[0,T]}) > \varepsilon\right] \le e^{-K\,N\,\varepsilon^2}$$

for all $\varepsilon > 0$ and $N \ge N_0\,\varepsilon^{-2}\,\exp\,(N_0\,\varepsilon^{-1/\alpha})$.

Here the constants K and N_0 depend on T, V, W, α and $\displaystyle\int_{\mathbb{R}^d} e^{a_0|x|^2}\, d\mu_0(x)$.

By Kantorovich-Rubinstein formulation again, this bound can be written as

$$\mathbb{P}\left[\sup_{[\varphi]_1\le 1}\Big(\frac{1}{N}\sum_{i=1}^N \varphi(X^i) - \int_{\mathcal{C}} \varphi(x)\, d\mu_{[0,T]}(x)\Big) > \varepsilon\right] \le e^{-K\,N\,\varepsilon^2}. \tag{7.12}$$

By projection at time t, it implies concentration inequalities for the time-marginals of the empirical measures similar to inequalities (7.1) and even (7.2). But above all it gives concentration estimates at the level of the whole paths. In return we impose some stronger condition on the required size of the sample (note however that by (7.6) one can also get less precise estimates valid for any number N of particles).

Assume for instance that one is interested in the behaviour of a point Y_t evolving according to (7.10). Then, from (7.12), one can derive error bounds in the approximation by $\displaystyle\frac{1}{N}\sum_{i=1}^N \varphi(X^i)$ of the expectation of quantities $\varphi(Y)$ which depend on the whole path, such as the distance $d(Y,A) = \inf\big\{|Y_t - y|; t \in [0,T], y \in A\big\}$ of the trajectory to a given set A in $\mathbb{R}^d$, which measures how close Y_t has been to A, or the maximal distance $\sup\big\{|Y_t - x|; t \in [0,T]\big\}$ to a given point x in the phase space $\mathbb{R}^d$: for instance, under the assumptions of Theorem 7.2, for any $T \ge 0$ and $\alpha \in (0,1/2)$ there exist some positive constants K and N_0 such that

$$\mathbb{P}\left[\Big|\mathbb{E}\,[d(Y,A)] - \frac{1}{N}\sum_{i=1}^N d(X^i, A)\Big| > \varepsilon\right] \le e^{-K\,N\,\varepsilon^2}$$

for any Borel set A in $\mathbb{R}^d$, $\varepsilon > 0$ and $N \geq N_0 \, \varepsilon^{-2} \, \exp\left(N_0 \, \varepsilon^{-1/\alpha}\right)$.

Theorem 7.2 will be proven in detail in Sections 7.3 and 7.4 along the following lines. Following A.-S. Sznitman [105] we proceed by coupling, by introducing a family of N identically distributed processes $Y^i = (Y^i_t)_{0 \leq t \leq T}$ solution to the (nonlinear) stochastic differential equations

$$\left\{ \begin{array}{rcl} dY^i_t & = & \sqrt{2}\, dB^i_t - \nabla V(Y^i_t)\, dt - \nabla W * \mu_t(Y^i_t)\, dt \\ Y^i_0 & = & X^i_0 \end{array} \right. \qquad 1 \leq i \leq N \, ;$$

here μ_t is the solution at time t to (6.16), but is also the law on $\mathbb{R}^d$ of any Y^i_t by Itô's formula, and, for each i, $B^i = (B^i_t)_{0 \leq t \leq T}$ is the Brownian motion driving the evolution of X^i. In particular the paths Y^i are close to the paths X^i and one can prove that there exists some constant C (depending only on T) such that

$$W_{1,[0,T]}(\mu_{[0,T]}, \hat{\mu}^N_{[0,T]}) \leq C \, W_{1,[0,T]}(\mu_{[0,T]}, \hat{\nu}^N_{[0,T]})$$

hold almost surely, where $\hat{\nu}^N_{[0,T]} := \dfrac{1}{N} \displaystyle\sum_{i=1}^N \delta_{Y^i}$; hence controlling the distance between $\mu_{[0,T]}$ and $\hat{\mu}^N_{[0,T]}$ reduces to the same issue with $\mu_{[0,T]}$ and $\hat{\nu}^N_{[0,T]}$.

But, by definition, the N processes Y^i for $1 \leq i \leq N$ are independent and distributed according to $\mu_{[0,T]}$. Then Theorem 7.1 ensures good concentration estimates for the empirical measure $\hat{\nu}^N_{[0,T]}$ around the common law $\mu_{[0,T]}$. In the end we obtain the bound

$$\mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]}, \hat{\mu}^N_{[0,T]}) > \varepsilon\right] \leq \mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]}, \hat{\nu}^N_{[0,T]}) > \frac{\varepsilon}{C}\right] \leq e^{-K \, N \, \varepsilon^2}$$

under some condition on ε and N.

The proof is actually an adaptation of the argument given in [22, Section 1.6] of estimates (7.2) for time-marginals. The current proof turns out to be simpler in the sense that it consists in fewer steps ; in return each of these steps is somehow more delicate : for instance, as we shall see in the following sections, the proof of Theorem 7.2 requires the computation of the metric entropy of some space of Hölder-continuous functions, and checking that the law of Y fulfills the assumptions of this theorem needs some strong integrability in Hölder norm on solutions to stochastic differential equations.

An adaptation of this proof leads to quantitative estimates on the phenomenon of **propagation of chaos**, namely, that any k particles drawn from the system tend to behave like independent and identically distributed variables. Indeed, considering, for instance for $k = 2$, the empirical measure on **pairs of paths** defined on $\mathcal{C} \times \mathcal{C}$ by

$$\hat{\mu}^{N,2}_{[0,T]} := \frac{1}{N(N-1)} \sum_{i \neq j} \delta_{(X^i, X^j)}$$

one can prove

Theorem 7.3. *With the same notation and assumptions as in Theorem 7.2, for all $T \geq 0$ and $\alpha \in (0, 1/2)$ there exist some positive constants K and N_0 such that*

$$\mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]} \otimes \mu_{[0,T]}, \hat{\mu}^{N,2}_{[0,T]}) > \varepsilon\right] \leq e^{-K\, N\, \varepsilon^2}$$

for all $\varepsilon > 0$ and $N \geq N_0\, \varepsilon^{-2} \exp(N_0\, \varepsilon^{-1/\alpha})$.

Here the constants K and N_0 depend on T, V, W, α and a finite square-exponential moment of μ_0, and $W_{1,[0,T]}$ stands for the Wasserstein distance of order 1 on the product space $\mathcal{C} \times \mathcal{C}$. The proof consists in writing the coupling argument for pairs of paths and comparing $\mu_{[0,T]} \otimes \mu_{[0,T]}$ and $\hat{\nu}^{N,2}_{[0,T]} := \dfrac{1}{N(N-1)} \displaystyle\sum_{i \neq j} \delta_{(Y^i, Y^j)}$ through $\dfrac{1}{N^2} \displaystyle\sum_{i,j} \delta_{(Y^i, Y^j)}$.

Let us finally note that it would be desirable to relax the assumptions made on the potentials V and W, in particular so as to include the interesting case of the cubic potential $W(z) = |z|^3/3$ on $\mathbb{R}$, which models the interaction among one-dimensional granular media (see [11]). It could also be interesting to consider the whole trajectories $(X_t)_{t \geq 0}$, and derive concentration bounds on functionals such as hitting times for instance.

Before turning to the proofs we briefly recall the **plan of the paper**. In the coming section we prove Theorem 7.1 for general $\mathcal{C}$-valued independent variables. The study of the particle system is addressed in the following two sections : in Section 3 we reduce our concentration issue on interacting particles to the same issue for independent variables by a coupling argument, whereas in Section 4 we check that we can apply our general concentration result to these independent variables ; with this in hand we can prove Theorem 7.2. An appendix is devoted to a general metric entropy estimate in a space of Hölder-continuous functions, which enters the proof of Theorem 7.1.

7.2 A preliminary result on independent variables

The aim of this section is to prove Theorem 7.1 for N independent and identically distributed random variables valued in $\mathcal{C}$. We have seen how this result, applied to the artificial processes $(Y^i_t)_t$, enters into the study of our interacting particle system.

The proof goes in three steps : truncation to a ball $\mathcal{B}^\alpha_R$ of $\mathcal{C}^\alpha$, compact for the topology induced by the uniform norm ; covering of $\mathcal{B}^\alpha_R$ and then of $\mathcal{P}(\mathcal{B}^\alpha_R)$ by small balls on which one develops Sanov's argument ; conclusion of the argument by optimizing the introduced parameters. Since it follows the lines of the argument given in [22, Section 2.1] in the finite dimensional case, in which μ is a measure on $\mathbb{R}^d$, we shall only sketch the argument, stressing only the bounds specific to our new framework. We refer to [22] for further details.

Step 1. Truncation. Given $R > 0$, to be chosen later on, we denote $\mathcal{B}^\alpha_R$ the ball $\{f \in \mathcal{C}^\alpha; \|f\|_\alpha \leq R\}$ of center 0 and radius R in $\mathcal{C}^\alpha$. This set $\mathcal{B}^\alpha_R$ is a compact subset of $\mathcal{C}$ for the topology induced by the uniform norm $\|\cdot\|_\infty$: indeed it is relatively compact in $\mathcal{C}$ by

Ascoli's theorem, and closed since if f in $\mathcal{C}$ is the uniform limit of a sequence $(f_n)_n$ in $\mathcal{C}^\alpha$, then $\|f\|_\alpha \leq \liminf_n \|f_n\|_\alpha$, and in particular f belongs to $\mathcal{B}_R^\alpha$ if so do the f_n.

Letting $\mathbf{1}_{\mathcal{B}_R^\alpha}$ be the indicator function of $\mathcal{B}_R^\alpha$, we truncate μ into a probability measure μ_R on the ball $\mathcal{B}_R^\alpha$, defined as

$$\mu_R := \frac{\mathbf{1}_{\mathcal{B}_R^\alpha}\,\mu}{\mu[\mathcal{B}_R^\alpha]}.$$

Note that $\mu[\mathcal{B}_R^\alpha]$ is positive for R larger than some R_0 depending only on a and $E_a := \displaystyle\int_{\mathcal{C}} e^{a\|x\|_\alpha^2}\,d\mu(x)$. In this step we reduce the concentration problem for $\mathcal{C}$ to the same issue for the compact ball $\mathcal{B}_R^\alpha$, by bounding the quantity $\mathbb{P}\big[W_p(\mu,\hat{\mu}^N) > \varepsilon\big]$ in terms of μ_R and the associated empirical measure.

This is achieved by the following coupling argument. Given independent random variables $(X^k)_{1\leq k\leq N}$ with law μ, and $(Y^k)_{1\leq k\leq N}$ with law μ_R, let

$$X_R^k := \begin{cases} X^k & \text{if } |X^k|_\alpha \leq R, \\ Y^k & \text{if } |X^k|_\alpha > R. \end{cases}$$

Since X^1 and X_R^1 have respective law μ and μ_R, we have, by definition of $W_{p,[0,T]}$ distance,

$$\begin{aligned} W_{p,[0,T]}(\mu,\mu_R)^p &\leq \mathbb{E}\|X^1 - X_R^1\|_\infty^p = \mathbb{E}\Big(\|X^1 - Y^1\|_\infty^p \mathbf{1}_{\|X^1\|_\alpha > R}\Big) \\ &\leq 2^{p-1}\mathbb{E}\big((\|X^1\|_\infty^p + R^p)\mathbf{1}_{\|X^1\|_\alpha > R}\big) \leq 2^p \int_{\{\|x\|_\alpha > R\}} \|x\|_\alpha^p\,d\mu(x). \end{aligned}$$

By monotonicity of the map $x \mapsto |x|^p\,e^{-a\,|x|^2}$ on $|x| \geq \sqrt{p/2a}$, this ensures the bound

$$W_{p,[0,T]}(\mu,\mu_R)^p \leq 2^p E_a R^p e^{-aR^2}\,, \qquad R \geq \max(R_0, \sqrt{p/2a}). \tag{7.13}$$

On the other hand, the empirical measures

$$\hat{\mu}^N := \frac{1}{N}\sum_{k=1}^N \delta_{X^k} \qquad \text{and} \qquad \hat{\mu}_R^N := \frac{1}{N}\sum_{k=1}^N \delta_{X_R^k}$$

satisfy

$$W_{p,[0,T]}(\hat{\mu}_R^N,\hat{\mu}^N)^p \leq \frac{1}{N}\sum_{k=1}^N |X_R^k - X^k|_\infty^p \leq \frac{1}{N}\sum_{k=1}^N 2^p\,\|X^k\|_\alpha^p\,\mathbf{1}_{|X^k|_\alpha > R}.$$

Then, proceeding as in [22, Section 2.1], we can use Chebyshev's exponential inequality and the independence of the variables X^k to obtain, given $a_1 < a$ and $1 \leq p < 2$, the existence of some constant R_1 such that

$$\mathbb{P}\big[W_{p,[0,T]}(\hat{\mu}_R^N,\hat{\mu}^N) > \varepsilon\big] \leq \exp\Big(-N\big(\theta\,\varepsilon^p - E_a e^{(a_1-a)R^2}\big)\Big) \tag{7.14}$$

for all $\theta > 0$ and $R \geq R_1 \theta^{\frac{1}{2-p}}$. From (7.13) and (7.14), and by triangular inequality for $W_{p,[0,T]}$, we obtain the bound

$$\mathbb{P}\left[W_{p,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \mathbb{P}\left[W_{p,[0,T]}(\mu_R, \hat{\mu}_R^N) > \eta\,\varepsilon - 2\,E_a^{1/p} R\, e^{-\frac{a}{p}R^2}\right] + \exp\Big(-N\big(\theta(1-\eta)^p\varepsilon^p - E_a\, e^{(a_1-a)\,R^2}\big)\Big); \quad (7.15)$$

here p is any real number in $[1,2)$, η in $(0,1)$, $\varepsilon, \theta > 0, a_1 < a$ and R is constrained to be larger than $R_2 \max(1, \theta^{\frac{1}{2-p}})$ for some constant R_2 depending only on E_a, a, a_1 and p.

In the case when $p = 2$, we obtain

$$\mathbb{P}\left[W_{2,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \mathbb{P}\left[W_{2,[0,T]}(\mu_R, \hat{\mu}_R^N) > \eta\,\varepsilon - 2\,E_a^{1/2} R\, e^{-\frac{a}{2}R^2}\right] + \exp\Big(-N\big(\frac{a_1}{2}(1-\eta)^2\varepsilon^2 - 2E_a^2\, e^{(a_1-a)\,R^2}\big)\Big). \quad (7.16)$$

Step 2. Sanov's argument on small balls. In view of (7.15) for $p < 2$ or (7.16) for $p = 2$, we now aim at bounding $\mathbb{P}[\hat{\mu}_R^N \in \mathcal{A}]$ where

$$\mathcal{A} := \Big\{\nu \in \mathcal{P}(\mathcal{B}_R^\alpha); \quad W_{p,[0,T]}(\nu, \mu_R) \geq \eta\,\varepsilon - 2\,E_a^{1/p} R\, e^{-\frac{a}{p}R^2}\Big\}.$$

For that purpose, reasoning as in [22, Section 2.1], we let $\delta > 0$ and cover $\mathcal{A}$ with $\mathcal{N}(\mathcal{A}, \delta)$ balls $(B_i)_{1 \leq i \leq \mathcal{N}(\mathcal{A},\delta)}$ with radius $\delta/2$ in $W_{p,[0,T]}$ distance. Then one can develop Sanov's argument on each of these compact and convex balls, to obtain the bound

$$\mathbb{P}[\hat{\mu}_R^N \in \mathcal{A}] \leq \mathbb{P}\Big[\hat{\mu}_R^N \in \bigcup_{i=1}^{\mathcal{N}(\mathcal{A},\delta)} B_i\Big] \leq \sum_{i=1}^{\mathcal{N}(\mathcal{A},\delta)} \mathbb{P}\left[\hat{\mu}_R^N \in B_i\right] \leq \sum_{i=1}^{\mathcal{N}(\mathcal{A},\delta)} \exp\Big(-N \inf_{\nu \in B_i} H(\nu|\mu_R)\Big). \quad (7.17)$$

Then one establishes an approximate $T_p(\lambda)$ inequality for μ_R : namely, for any $\lambda_1 < \lambda$ there exists K_1 such that

$$H(\nu, \mu_R) \geq \frac{\lambda_1}{2}\, W_{p,[0,T]}(\nu, \mu_R)^2 - K_1\, R^2\, e^{-a\,R^2}$$

for all measure ν on $\mathcal{B}_R^\alpha$. With this inequality in hand, given $1 \leq p < 2$ and $\lambda_2 < \lambda_1 < \lambda$, one deduces from (7.17) the existence of some positive constants δ_1, η_1 and K_1 such that

$$\mathbb{P}\left[W_{p,[0,T]}(\mu_R, \hat{\mu}_R^N) > \eta\,\varepsilon - 2\,E_a^{1/p} R e^{-\frac{a}{p}R^2}\right] \leq \mathcal{N}(\mathcal{A}, \delta)\, \exp\left(-N\Big(\frac{\lambda_2}{2}\,\varepsilon^2 - K_1 R^2 e^{-aR^2}\Big)\right) \quad (7.18)$$

where we have chosen $\delta := \delta_1 \varepsilon$ and $\eta := \eta_1$.

In the case when $p = 2$, we do not choose η at this stage, and simply obtain

$$\mathbb{P}\left[W_{2,[0,T]}(\mu_R, \hat{\mu}_R^N) > \eta\,\varepsilon - 2\,E_a^{1/2} R\, e^{-\frac{a}{2}R^2}\right] \leq \mathcal{N}(\mathcal{A}, \delta)\, \exp\left(-N\Big(\frac{\lambda_2}{2}\,\eta^2\varepsilon^2 - K_1 R^2 e^{-aR^2}\Big)\right)$$

where $\delta := \delta_1 \varepsilon$.

Then, since $\mathcal{A}$ is a subset of $\mathcal{P}(\mathcal{B}_R^\alpha)$, Theorem 7.10 in the Appendix enables to bound $\mathcal{N}(\mathcal{A}, \delta)$ with $\delta = \delta_1 \varepsilon$ by

$$\exp\Big(K_2 (R\,\varepsilon^{-1})^d\, 3^{K_2 (R\,\varepsilon^{-1})^{1/\alpha}}\ \ln\big(\max(1, K_2\, R\,\varepsilon^{-1})\big)\Big) \tag{7.19}$$

for some constant K_2 depending neither on ε nor on R.

Remark 7.4. The order of magnitude of this covering number in an infinite-dimensional setting constitutes a main change by comparison with the finite-dimensional setting of [22], and will influence the final condition on the size N of the sample.

Step 3. Conclusion of the argument. We first focus on the case when $p \in [1, 2)$. Collecting estimates (7.15), (7.18) and (7.19), we obtain, given $\lambda_2 < \lambda$ and $a_1 < a$, the existence of some positive constants K_1, K_2, K_3 and R_3 depending on $E_a, a, a_1, \alpha, \lambda$ and λ_2 such that

$$\begin{aligned}
&\mathbb{P}\left[W_{p,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \\
&\quad \leq \exp\bigg(K_2 (R\,\varepsilon^{-1})^d\, 3^{K_2 (R\,\varepsilon^{-1})^{1/\alpha}}\ \ln\big(\max(1, K_2\, R\,\varepsilon^{-1})\big) - N\Big(\frac{\lambda_2}{2}\,\varepsilon^2 - K_1 R^2 e^{-\alpha R^2}\Big)\bigg) \\
&\qquad\qquad + \exp\Big(-N\big(K_3\,\theta\,\varepsilon^p - K_4 e^{(a_1 - a)\,R^2}\big)\Big)
\end{aligned} \tag{7.20}$$

for all $\varepsilon, \theta > 0$ and $R \geq R_3 \max(1, \theta^{\frac{1}{2-p}})$, and for some constant $K_4 = K_4(\theta, a_1)$.

Then let $\lambda_3 < \lambda_2$. One can prove that the first term in the right-hand side in (7.20) is bounded by $\exp\left(-\dfrac{\lambda_3}{2}\, N\,\varepsilon^2\right)$ provided

$$R^2 \geq R_4 \max(1, \varepsilon^2, \ln(\varepsilon^{-2}))\,, \quad N\varepsilon^2 \geq K_5\, 3^{K_6 (R\varepsilon^{-1})^{1/\alpha}}$$

for some positive constants R_4, K_5 and K_6 depending also on λ_3. Moreover, for

$$\theta = \frac{\varepsilon^{2-p}\,\lambda_3}{2\,K_3},$$

also the second term in the right-hand side in (7.20) is bounded by $\exp\left(-\dfrac{\lambda_3}{2}\, N\,\varepsilon^2\right)$ as soon as $R^2 \geq R_5 \max(1, \ln(\varepsilon^{-2}))$, for some constant R_5 depending on λ_3.

To sum up, given $\lambda_3 < \lambda$, there exist some positive constants A, B and C such that

$$\mathbb{P}\left[W_p(\mu, \hat{\mu}^N) > \varepsilon\right] \leq 2\,\exp\Big(-\frac{\lambda_3}{2}\, N\,\varepsilon^2\Big)$$

as soon as

$$R^2 \geq A \max\big(1, \varepsilon^2, \ln(\varepsilon^{-2})\big), \qquad N\varepsilon^2 \geq B\,\exp\big(C (R\,\varepsilon^{-1})^{1/\alpha}\big)\,. \tag{7.21}$$

Letting $R = \varepsilon \left(\dfrac{1}{C} \ln \dfrac{N\varepsilon^2}{B} \right)^{\alpha}$ if $\varepsilon \in (0,1)$ and $R = \sqrt{A}\,\varepsilon$ otherwise, and $\alpha' < \alpha$, both conditions in (7.21) hold true as soon as $N \geq N_0\,\varepsilon^{-2} \exp(N_0\,\varepsilon^{-1/\alpha'})$ for some constant N_0 depending on $E_a, a, \lambda, \lambda_3, \alpha$ and α'. Finally, given $\lambda' < \lambda_3 < \lambda$, this condition ensures that

$$\mathbb{P}\left[W_{p,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \exp\Big(- \frac{\lambda'}{2}\, N\, \varepsilon^2 \Big),$$

possibly for some larger N_0. This concludes the argument in the case when $p \in [1,2)$.

In the case when $p = 2$, given $0 < \eta < 1, \lambda_3 < \lambda_2$ and $a_2 < a_1$, the same condition on N and ε (for some N_0) is sufficient for the bound

$$\mathbb{P}\left[W_{2,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \exp\left(-\frac{\lambda_3}{2}\,\eta^2\, N\, \varepsilon^2 \right) + \exp\left(-\frac{a_2}{2} (1-\eta)^2\, N\, \varepsilon^2 \right)$$

to hold (by (7.16)). One optimizes this bound by letting

$$a_2 = a\,\frac{\lambda_3}{\lambda}\,(\in [0,a)) \quad \text{and} \quad \eta = \frac{\sqrt{a_2}}{\sqrt{a_2} + \sqrt{\lambda_3}},$$

which gives

$$\mathbb{P}\big[W_{2,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\big] \leq 2\, \exp\left(-\frac{\lambda_3}{2} \frac{a}{(\sqrt{a} + \sqrt{\lambda})^2}\, N\, \varepsilon^2 \right).$$

Finally, given $\lambda' < \lambda_3 < \lambda$, there exists some (possibly larger) N_0 such that

$$\mathbb{P}\big[W_{2,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\big] \leq \exp\left(-\frac{\lambda'}{2} \frac{a}{(\sqrt{a} + \sqrt{\lambda})^2}\, N\, \varepsilon^2 \right).$$

for any $\varepsilon > 0$ and $N \geq N_0\,\varepsilon^{-2} \exp(N_0\,\varepsilon^{-1/\alpha'})$. This concludes the proof of Theorem 7.1 in this second and last case. □

7.3 Coupling

Here we begin the proof of Theorem 7.2 on the behaviour of our interacting particle system whose size tends to infinity.

We recall that we are given N independent variables X_0^i in $\mathbb{R}^d$, with common law μ_0, and N independent Brownian motions $B^i = (B_t^i)_{0 \leq t \leq T}$ in $\mathbb{R}^d$, and we consider the solutions $X^i = (X_t^i)_{0 \leq t \leq T}$ to the coupled stochastic differential equations

$$dX_t^i = \sqrt{2}\, dB_t^i - \nabla V(X_t^i)\, dt - \frac{1}{N} \sum_{j=1}^{N} \nabla W(X_t^i - X_t^j)\, dt \qquad 1 \leq i \leq N.$$

We also let $\mu_{[0,T]}$ be the law of the process $Y = (Y_t)_{0 \leq t \leq T}$ defined by

$$dY_t = \sqrt{2}\, dB_t - \nabla V(Y_t)\, dt - \nabla W * \mu_t(Y_t)\, dt$$

and starting at some Y_0 drawn according to μ_0 ; here $B = (B_t)_{0 \leq t \leq T}$ also is a Brownian motion and μ_t is the law of Y_t, that is, the time-marginal of $\mu_{[0,T]}$ at time t.

We want to compare this law $\mu_{[0,T]}$ and the empirical measure of the paths

$$\hat{\mu}^N_{[0,T]} = \frac{1}{N} \sum_{i=1}^N \delta_{X^i}.$$

For this purpose we introduce the family of N independent processes $Y^i = (Y^i_t)_{0 \leq t \leq T}$ defined by

$$dY^i_t = \sqrt{2}\, dB^i_t - \nabla V(Y^i_t)\, dt - \nabla W * \mu_t(Y^i_t)\, dt \qquad 1 \leq i \leq N \tag{7.22}$$

for the same Brownian motions B^i, and such that $Y^i_0 = X^i_0$ initially. We let

$$\hat{\nu}^N_{[0,T]} := \frac{1}{N} \sum_{i=1}^N \delta_{Y^i} \tag{7.23}$$

and in this section we reduce the problem to measuring the distance between $\mu_{[0,T]}$ and $\hat{\nu}^N_{[0,T]}$.

Indeed we prove

Proposition 7.5. *In the above notation and under the assumptions*

$$\beta\, I \leq D^2 V(x)\,, \quad \gamma\, I \leq D^2 W(x) \leq \gamma'\, I\,, \qquad x \in \mathbb{R}^d$$

on V and W, where β, γ and γ' are real numbers, for any $T \geq 0$ there exists some constant C depending only on β, γ, γ' and T such that

$$W_{1,[0,T]}(\mu_{[0,T]}, \hat{\mu}^N_{[0,T]}) \leq C\, W_{1,[0,T]}(\mu_{[0,T]}, \hat{\nu}^N_{[0,T]})$$

almost surely.

Proof. We first follow the lines of the proof of [22, Proposition 5.1], but in the end we want an estimate on the trajectories as a whole. Since for each i both processes $(X^i_t)_t$ and $(Y^i_t)_t$ are driven by the same Brownian motion B^i, the process $X^i - Y^i$ satisfies the equation

$$d(X^i_t - Y^i_t) = -\big(\nabla V(X^i_t) - \nabla V(Y^i_t)\big)\, dt - \big(\nabla W * \hat{\mu}^N_t(X^i_t) - \nabla W * \mu_t(Y^i_t)\big)\, dt.$$

In particular, letting $u \cdot v$ denote the scalar product of two vectors u and v in $\mathbb{R}^d$,

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}|X^i_t - Y^i_t|^2 = &- \big(\nabla V(X^i_t) - \nabla V(Y^i_t)\big) \cdot (X^i_t - Y^i_t) \\ &- \big(\nabla W * \hat{\mu}^N_t(X^i_t) - \nabla W * \mu_t(Y^i_t)\big) \cdot (X^i_t - Y^i_t). \end{aligned} \tag{7.24}$$

We decompose the last term according to

$$\begin{aligned} &\nabla W * \hat{\mu}^N_t(X^i_t) - \nabla W * \mu_t(Y^i_t) = \\ &\qquad (\nabla W * \hat{\mu}^N_t - \nabla W * \mu_t)(X^i_t) + \big(\nabla W * \mu_t(X^i_t) - \nabla W * \mu_t(Y^i_t)\big). \end{aligned}$$

By our assumption on D^2W, the map $\nabla W(X_t^i - \cdot\,)$ is Γ-Lipschitz with $\Gamma := \max(|\gamma|, |\gamma'|)$. Consequently, by the Kantorovich-Rubinstein dual formulation (7.5) of $W_{1,\tau}$,

$$\Big|\nabla W * (\hat{\mu}_t^N - \mu_t)(X_t^i)\Big| = \Big|\int_{\mathbb{R}^d} \nabla W(X_t^i - y)\, d(\hat{\mu}_t^N - \mu_t)(y)\Big| \leq \Gamma\, W_{1,\tau}(\hat{\mu}_t^N, \mu_t). \tag{7.25}$$

Then, in view of our convexity assumptions on V and W, (7.24) and (7.25) imply

$$\frac{1}{2}\frac{d}{dt}|X_t^i - Y_t^i|^2 \leq -(\gamma+\beta)\,|X_t^i - Y_t^i|^2 + \Gamma\, W_{1,\tau}(\hat{\mu}_t^N, \mu_t)\,|X_t^i - Y_t^i|.$$

In particular, by Gronwall's lemma,

$$|X_t^i - Y_t^i| \leq \Gamma \int_0^t e^{-(\beta+\gamma)(t-u)}\, W_{1,\tau}(\hat{\mu}_u^N, \mu_u)\, du$$

since initially $X_0^i = Y_0^i$. Consequently, by convexity of the $W_{1,[0,t]}$ distance,

$$\begin{aligned} W_{1,[0,t]}(\hat{\mu}_{[0,t]}^N, \hat{\nu}_{[0,t]}^N) &\leq \frac{1}{N}\sum_{i=1}^N \sup_{0\leq s\leq t} |X_s^i - Y_s^i| \\ &\leq \frac{1}{N}\sum_{i=1}^N \sup_{0\leq s\leq t} \Gamma \int_0^s e^{-(\beta+\gamma)(s-u)} W_{1,\tau}(\hat{\mu}_u^N, \mu_u)\, du \\ &\leq \Gamma e^{|\beta+\gamma| T} \int_0^t W_{1,\tau}(\hat{\mu}_u^N, \mu_u)\, du \end{aligned} \tag{7.26}$$

for all $0 \leq t \leq T$. But

$$W_{1,\tau}(\hat{\mu}_u^N, \mu_u) \leq W_{1,[0,u]}(\hat{\mu}_{[0,u]}^N, \mu_{[0,u]}) \leq W_{1,[0,u]}(\hat{\mu}_{[0,u]}^N, \hat{\nu}_{[0,u]}^N) + W_{1,[0,u]}(\hat{\nu}_{[0,u]}^N, \mu_{[0,u]})$$

by the projection relation (7.3) and triangular inequality for $W_{1,[0,u]}$, so by Gronwall's lemma again

$$W_{1,[0,t]}(\hat{\mu}_{[0,t]}^N, \hat{\nu}_{[0,t]}^N) \leq \Gamma e^{|\beta+\gamma| T} \int_0^t \exp\Big(\Gamma e^{|\beta+\gamma| T}(t-u)\Big)\, W_{1,[0,u]}(\hat{\nu}_{[0,u]}^N, \mu_{[0,u]})\, du$$

for all $0 \leq t \leq T$. But

$$W_{1,[0,u]}(\hat{\nu}_{[0,u]}^N, \mu_{[0,u]}) \leq W_{1,[0,t]}(\hat{\nu}_{[0,t]}^N, \mu_{[0,t]})$$

for all $0 \leq u \leq t$ since $\hat{\nu}_{[0,u]}^N$ and $\mu_{[0,u]}$ are the respective image measures of $\hat{\nu}_{[0,t]}^N$ and $\mu_{[0,t]}$ by the 1-Lipschitz map defined from $\mathcal{C}([0,t], \mathbb{R}^d)$ into $\mathcal{C}([0,u], \mathbb{R}^d)$ as the restriction to $[0,u]$. From this we obtain the bound

$$W_{1,[0,t]}(\hat{\mu}_{[0,t]}^N, \hat{\nu}_{[0,t]}^N) \leq C\, W_{1,[0,t]}(\hat{\nu}_{[0,t]}^N, \mu_{[0,t]})$$

for some constant C depending only on T, β, γ and γ'. This concludes the argument by triangular inequality. $\square$

Remark 7.6. If moreover $\beta + \gamma > \Gamma$ where again $\Gamma := \max(|\gamma|, |\gamma'|)$, then we can let C be $(\beta+\gamma)\,(\beta+\gamma-\Gamma)^{-1}$ in Proposition 7.5, independently of T. Indeed, if $\beta+\gamma > 0$, then (7.26) leads to

$$W_{1,[0,t]}(\hat{\mu}^N_{[0,t]}, \hat{\nu}^N_{[0,t]}) \leq \Gamma \sup_{0\leq s\leq t} \int_0^t e^{-(\beta+\gamma)(s-u)}\, du \sup_{0\leq u\leq t} W_{1,\tau}(\hat{\mu}^N_u, \mu_u) \leq \frac{\Gamma}{\beta+\gamma}\, W_{1,[0,t]}(\hat{\mu}^N_{[0,t]}, \mu_{[0,t]})$$

and by triangular inequality

$$W_{1,[0,t]}(\hat{\mu}^N_{[0,t]}, \mu_{[0,t]}) \leq \frac{\beta+\gamma}{\beta+\gamma-\Gamma}\, W_{1,[0,t]}(\hat{\nu}^N_{[0,t]}, \mu_{[0,t]})$$

provided $\beta + \gamma > \Gamma$.

This is reminiscent of the fact that, under some convexity assumptions on V and W, such as $\beta > 0$, $\beta + 2\,\gamma > 0$, it has been proven in [74, 36, 37] that the time-marginal μ_t of the measure $\mu_{[0,t]}$ converges, as t goes to infinity, to the stationary solution to the limit equation (7.9). One can also prove in this context that (in expectation) observables of the particle system are bounded in time.

Hence, under this kind of assumptions, one could hope for some uniform in time constants in this coupling argument : that was obtained in [22, Proposition 5.1] for the time-marginals, and here for the whole processes. However, contrary to [22] where this property was used to approach the stationary solution by coupling together estimates of concentration of the empirical measure (as N goes to infinity) with estimates of convergence to equilibrium (as t goes to infinity), in this work we are concerned with finite time intervals only, and shall not use this specific property in the sequel.

7.4 Conclusion of the argument

In the previous section we have reduced the issue of measuring the distance between $\mu_{[0,T]}$ and $\hat{\mu}^N_{[0,T]}$ to measuring the distance between $\mu_{[0,T]}$ and the empirical measure $\hat{\nu}^N_{[0,T]}$ of N independent random variables drawn according to $\mu_{[0,T]}$.

We now solve the latter issue by proving that the measure $\mu_{[0,T]}$ fulfills the hypotheses of Theorem 7.1 with $p = 1$, namely, that there exist some $\alpha \in (0, 1]$ and $a > 0$ such that

$$\int_{\mathcal{C}} e^{a\|x\|^2_\alpha}\, d\mu_{[0,T]}(x) := \mathbb{E}\exp(a\|Y_\cdot\|^2_\alpha) < +\infty.$$

Here again $\mathcal{C}$ stands for $\mathcal{C}([0, T], \mathbb{R}^d)$, $\|f\|_\alpha$ for the Hölder norm of f on $[0, T]$, and $Y = (Y_t)_{0\leq t\leq T}$ is the solution to the stochastic differential equation

$$dY_t = \sqrt{2}\, dB_t - \nabla V(Y_t)\, dt - \nabla W * \mu_t(Y_t)\, dt \tag{7.27}$$

starting at Y_0 drawn according to μ_0, where μ_t is the law of Y_t.

Proposition 7.7. *Let μ_0 be a Borel probability measure on $\mathbb{R}^d$ admitting a finite square-exponential moment, and let Y_0 be drawn according to μ_0. Given $T \geq 0$, V and W satisfying hypotheses* (7.11)*, let Y be the solution to* (7.27) *starting at Y_0. Then, for any $\alpha \in (0, 1/2)$, there exists $a > 0$, depending on μ_0 only through a finite square-exponential moment, such that $\mathbb{E} \exp(a\|Y_\cdot\|_\alpha^2)$ be finite.*

Assuming this result for the moment we can now conclude the **proof of Theorem 7.2**. Let indeed α be given in $(0, 1/2)$, and $\alpha_0 \in (\alpha, 1/2)$. Then, by Proposition 7.7 and Theorem 7.1, applied with $\alpha = \alpha_0$ and $\alpha' = \alpha$, there exist some constants $\tilde{K}$ and $\tilde{N}_0$, depending on α_0, α, T and a square-exponential moment of μ_0, such that

$$\mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]}, \hat{\nu}^N_{[0,T]}) > \tilde{\varepsilon}\right] \leq e^{-\tilde{K}N\tilde{\varepsilon}^2}$$

for any $\tilde{\varepsilon} > 0$ and $N \geq \tilde{N}_0\,\tilde{\varepsilon}^{-2}\,\exp(\tilde{N}_0\,\tilde{\varepsilon}^{-1/\alpha})$, where $\hat{\nu}^N_{[0,T]}$ is defined by (7.22) and (7.23). Then, by Proposition 7.5, there exist some constants C, depending only on T, and then K and N_0, depending on α_0, α, T and a finite square-exponential moment of μ_0, such that

$$\mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]}, \hat{\mu}^N_{[0,T]}) > \varepsilon\right] \leq \mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]}, \hat{\nu}_{[0,T]}) > \varepsilon/C\right] \leq e^{-KN\varepsilon^2}$$

for any $\varepsilon > 0$ and $N \geq N_0\,\varepsilon^{-2}\,\exp(N_0\,\varepsilon^{-1/\alpha})$. This concludes the argument. □

Proof of Proposition 7.7. It is necessary and sufficient to prove that there exist positive constants a_1 and a_2 such that $\mathbb{E}\exp(a_1\|Y\|_\infty^2)$ and $\mathbb{E}\exp(a_2[Y]_\alpha^2)$ be finite, where $[\,\cdot\,]_\alpha$ stands for the Hölder seminorm defined in Section 7.1.2.

1. We start with the expectation in uniform norm. For this we first note that, according to [22, Proposition 3.1], there exist some positive constants M and $\overline{a}$, depending on μ_0 only through a finite square-exponential moment, such that $\sup_{0\leq t\leq T} \mathbb{E}\,|Y_t|^2$ and $\sup_{0\leq t\leq T} \mathbb{E}\exp(\overline{a}|Y_t|^2)$ be finite and bounded by M.

Then we let b be some smooth function on $[0,T]$, to be chosen later on, and we let $Z_t = \exp(b(t)\,|Y_t|^2)$. We want to prove that $\mathbb{E}\sup_{0\leq t\leq T} Z_t$ is finite for some positive function b. By Itô's formula,

$$Z_t = Z_0 + M_t + \int_0^t \left[b'(s)|Y_s|^2 + 2\,d\,b(s) + 4\,b(s)^2\,|Y_s|^2 - 2\,b(s)\,Y_s\cdot(\nabla V(Y_s) + \nabla W * \mu_s(Y_s))\right] Z_s\,ds$$

where $(M_t)_{0\leq t\leq T}$ is the martingale defined as

$$M_t = 2\sqrt{2}\int_0^t b(s)\,Z_s\,Y_s\,\cdot\,dB_s.$$

But $D^2V(x) \geq \beta\,I$ for all $x \in \mathbb{R}^d$, so for any $\delta > 0$ and $y \in \mathbb{R}^d$ we have

$$-y\,\cdot\,\nabla V(y) \leq (\delta - \beta)|y|^2 + \frac{|\nabla V(0)|^2}{4\delta}.$$

Furthermore ∇W is Γ-Lipschitz and $\nabla W(0) = 0$, so

$$\begin{aligned}
-2\,y\cdot\nabla W * \mu_s(y) &= -2\int_{\mathbb{R}^d} y\cdot\nabla W(y-z)\,d\mu_s(z)\\
&\le 2\,\Gamma\int_{\mathbb{R}^d}|y|\,|y-z|\,d\mu_s(z) \le 3\,\Gamma\,|y|^2 + \Gamma\int_{\mathbb{R}^d}|z|^2\,d\mu_s(z).
\end{aligned}$$

But $\displaystyle\int_{\mathbb{R}^d}|z|^2\,d\mu_s(z) = \mathbb{E}\,|Y_s|^2$ is bounded by M on $[0,T]$, so collecting all terms together, we obtain

$$Z_t \le Z_0 + M_t + \int_0^t \left[C(s) + D(s)\,|Y_s|^2\right] Z_s\,ds$$

where

$$C(s) = \Big(2\,d + \Gamma\,M + \frac{|\nabla V(0)|^2}{2\delta}\Big)\,b(s), \qquad D(s) = b'(s) + 4\,b(s)^2 + \big(2\,(\delta-\beta) + 3\,\Gamma\big)\,b(s).$$

Then, given $\delta > 0$ such that $c := 2\,(\delta-\beta) + 3\,\Gamma$ be positive, we let $b(s)$ such that $D(s) \equiv 0$, that is, let

$$b(s) = e^{-cs}\left(b(0)^{-1} + 4\,c^{-1}(1-e^{-cs})\right)^{-1} \tag{7.28}$$

for some $b(0)$ to be chosen later on. In particular b is a nonincreasing continuous positive function on $[0,+\infty)$, and, for this function b, Z_t almost surely satisfies the inequality

$$Z_t \le Z_0 + M_t + C(0)\int_0^t Z_s\,ds.$$

In particular

$$\mathbb{E}\sup_{0\le t\le T} Z_t \le \mathbb{E}\,Z_0 + \mathbb{E}\sup_{0\le t\le T} M_t + C(0)\int_0^T \mathbb{E}\,Z_s\,ds. \tag{7.29}$$

But, by Cauchy-Schwarz' and Doob's inequalities,

$$\Big(\mathbb{E}\sup_{0\le t\le T} M_t\Big)^2 \le \mathbb{E}\sup_{0\le t\le T}|M_t|^2 \le 2\sup_{0\le t\le T}\mathbb{E}\,|M_t|^2.$$

Then, by Itô's formula again,

$$\begin{aligned}
\mathbb{E}|M_t|^2 &= 8\int_0^t b(s)^2\,\mathbb{E}\left[Z_s^2\,|Y_s|^2\right] ds\\
&\le 8\,b(0)\int_0^t \mathbb{E}\left[b(s)\,|Y_s|^2\,\exp(2\,b(s)\,|Y_s|^2)\right] ds\\
&\le 8\,b(0)\int_0^t \mathbb{E}\,\exp(3\,b(0)\,|Y_s|^2)\,ds.
\end{aligned}$$

Choosing $b(0) \le \overline{a}/3$, this ensures that $\displaystyle\sup_{0\le t\le T}\mathbb{E}|M_t|^2$, whence $\displaystyle\mathbb{E}\sup_{0\le t\le T} M_t$, is finite.

Since, for this $b(0)$, $\sup_{0\leq t\leq T} \mathbb{E}\, Z_t$ also is finite, it follows from (7.29) that so is $\mathbb{E} \sup_{0\leq t\leq T} Z_t$, which concludes the argument for the expectation in uniform norm with $a_1 = b(T)$.

2. We now turn to the expectation in Hölder seminorm. For this purpose we simply write the solution as

$$Y_t = Y_0 + B_t - \int_0^t \big(\nabla V(Y_s) + \nabla W * \mu_s(Y_s)\big)\, ds$$

so that

$$[Y]_\alpha \leq [B]_\alpha + \Big[\int_0^\cdot \big(\nabla V(Y_s) + \nabla W * \mu_s(Y_s)\big)\, ds\Big]_\alpha$$

almost surely ; here Y and B stand as before for the map $t \mapsto Y_t$ and $t \mapsto B_t$ respectively, and $\int_0^\cdot \varphi(s)\, ds$ is an antiderivative of φ. Hence, by Cauchy-Schwarz' inequality,

$$\mathbb{E}\exp(a_2[Y]_\alpha^2) \leq \big(\mathbb{E}\exp(4\, a_2[B]_\alpha^2)\big)^{1/2}\Big(\mathbb{E}\exp\Big(4\, a_2\Big[\int_0^\cdot (\nabla V(Y_s) + \nabla W * \mu_s(Y_s))\, ds\Big]_\alpha^2\Big)\Big)^{1/2}.$$

But, on one hand, $\mathbb{E}\exp\big(4\, a_2[B]_\alpha^2\big)$ is finite for a_2 small enough (see [49, Theorem 1.3.2] for instance, with $E = \mathcal{C}$ and $N(f) = [f]_\alpha$). On the other hand, by assumption (7.11), ∇V and ∇W are respectively B and Γ-Lipschitz with $\mathrm{B} := \max(|\beta|, |\beta'|)$ and $\Gamma := \max(|\gamma|, |\gamma'|)$, so there exists some constant A such that

$$\big|\nabla V(y) + \nabla W * \mu_s(y)\big| \leq \mathrm{A} + (\mathrm{B} + \Gamma)|y|$$

for all $y \in \mathbb{R}^d$ and $s \in [0, T]$. In particular

$$\begin{aligned} \Big[\int_0^\cdot \big(\nabla V(Y_s) + \nabla W * \mu_s(Y_s)\big)\, ds\Big]_\alpha &\leq \sup_{0\leq s,t\leq T} \frac{1}{|t-s|^\alpha}\int_s^t \big(\mathrm{A} + (\mathrm{B} + \Gamma)|Y_u|\big)\, du \\ &\leq T^{1-\alpha}\big(\mathrm{A} + (\mathrm{B} + \Gamma)\|Y\|_\infty\big) \end{aligned}$$

almost surely, and

$$\begin{aligned} &\mathbb{E}\exp\Big(4\, a_2\Big[\int_0^\cdot \big(\nabla V(Y_s) + \nabla W * \mu_s(Y_s)\big)\, ds\Big]_\alpha^2\Big) \\ &\qquad\qquad \leq \exp\big(8\, a_2\, T^{2-2\alpha}\mathrm{A}^2\big)\, \mathbb{E}\exp\Big(8\, a_2\, T^{2-2\alpha}(\mathrm{B} + \Gamma)^2\|Y\|_\infty^2\Big) \end{aligned}$$

which by step 1 is finite as soon as $8\, a_2\, T^{2-2\alpha}(\mathrm{B} + \Gamma)^2 \leq a_1$.

On the whole, $\mathbb{E}\exp(a_2[Y]_\alpha^2)$ is indeed finite for a_2 small enough, depending on μ_0 only through a finite square-exponential moment, which concludes the argument. $\square$

7.5 Appendix : metric entropy of a Hölder space

The aim of this appendix is to establish the bound (7.19) used in the covering argument in the proof of Theorem 7.1, which amounts to studying the metric entropy of a Hölder space and of some related space of probability measures.

In the notation introduced in Sections 7.1.1 and 7.1.2, it is a consequence of Ascoli's theorem that the closed ball $\mathcal{B}^\alpha_R := \mathcal{B}^\alpha_R([0,T],\mathbb{R}^d) = \{f \in \mathcal{C}^\alpha; \|f\|_\alpha \leq R\}$ of center 0 and radius R in $\mathcal{C}^\alpha$ is a compact metric space for the metric defined by the uniform norm.

Here we aim at estimating by how many balls of given radius $r < R$ and centered in $\mathcal{B}^\alpha_R$ the compact metric space $\mathcal{B}^\alpha_R$ can be covered. We note that for $r \geq R$ the sole ball $\{f \in \mathcal{B}^\alpha_R; \|f\|_\infty \leq r\}$ covers $\mathcal{B}^\alpha_R$.

Notation : Given $r > 0$, the *covering number* $\mathcal{N}(E,r)$ of a compact metric space (E,d) is defined as the infimum of the integers n such that E can be covered by n balls centered in E and of radius r in d metric. Then we have the following result which gives some lower and upper bounds on the covering number $\mathcal{N}(\mathcal{B}^\alpha_R, r)$ and in our case makes more precise the bounds given for instance in [72] or [110] :

Theorem 7.8. *Given some integer number $d \geq 1$, some positive numbers T, R, r and α with $r < R$ and $\alpha \leq 1$, the covering number $\mathcal{N}(\mathcal{B}^\alpha_R, r)$ of $\mathcal{B}^\alpha_R$, equipped with the uniform norm, satisfies*

$$\mathcal{N}(\mathcal{B}^\alpha_R, r) \leq \Big(10\sqrt{d}\,\frac{R}{r}\Big)^d \, 3^{5^{\frac{1}{\alpha}}\, d^{1+\frac{1}{2\alpha}}\, T\,(\frac{R}{r})^{\frac{1}{\alpha}}}.$$

If moreover, for instance, $r \leq \dfrac{T^\alpha}{4\,T^\alpha + 4} R$, then

$$\mathcal{N}(\mathcal{B}^\alpha_R, r) \;\geq\; \Big(\frac{\sqrt{d}}{4}\,\frac{R}{r}\Big)^d \, 2^{2^{-\frac{1}{\alpha}}\, d^{1+\frac{1}{2\alpha}}\, T\,(\frac{R}{r})^{\frac{1}{\alpha}}}.$$

The lower bound ensures that the upper bound, from which depends the condition on the size of the sample imposed in Theorems 7.1 and hence 7.2, has the good order of growth in R/r.

Proof. 1. We start by establishing the upper bound.

1. 1. We first consider the case when $d = 1$.

Given J and K some integers larger or equal to 1, we let $\tau = \dfrac{T}{J}$ and $\eta = \dfrac{R}{K}$, and then

$$\begin{array}{rclrl} t_j & = & (j - \frac{1}{2})\,\tau\,, \quad j \in \mathbb{N}\,, & 1 & \leq j \leq J, \\ y_k & = & (k - \frac{1}{2})\,\eta\,, \quad k \in \mathbb{N}\,, & -K+1 & \leq k \leq K. \end{array}$$

Then we cover the rectangle $[0,T] \times [-R,+R]$ in $\mathbb{R}_t \times \mathbb{R}_y$, which contains the graph of all functions in $\mathcal{B}^\alpha_R([0,T],\mathbb{R})$, by a lattice with step τ in $t-$axis and η in $y-$axis.

Then let f be a given function in $\mathcal{B}_R^\alpha([0,T],\mathbb{R})$. Since the intervals $[y_k - \frac{\eta}{2}\,, y_k + \frac{\eta}{2}]$ cover the interval $[-R,+R]$, for every integer $j \in [1,J]$ there exists some integer $k(j) \in [-K+1,+K]$ such that

$$|f(t_j) - y_{k(j)}| \leq \frac{\eta}{2}.$$

In particular

$$|y_{k(j+1)} - y_{k(j)}| \leq \frac{\eta}{2} + |f(t_{j+1}) - f(t_j)| + \frac{\eta}{2} \leq \eta + R\,|t_{j+1} - t_j|^\alpha \leq \eta + R\,\tau^\alpha < 2\eta$$

if we suppose $KT^\alpha < J^\alpha$. But since the y_k take values which are regularly distant of η, it follows that more precisely

$$|y_{k(j+1)} - y_{k(j)}| \leq \eta.$$

From this map $k : [1,J] \cap \mathbb{N} \to [-K+1, K] \cap \mathbb{N}$, we define the function $f_k : [0,T] \to [-R,+R]$ affine on each interval of the subdivision $(0, t_1, \cdots, t_J, T)$ and such that

$$\begin{array}{rcl} f_k(0) & = & f_k(t_1), \\ f_k(t_j) & = & y_{k(j)}, \quad 1 \leq j \leq J \\ f_k(T) & = & f_k(t_J). \end{array}$$

In particular we note that this function f_k is Lipschitz with

$$\sup_{0 \leq t,s \leq T} \frac{|f_k(t) - f_k(s)|}{|t-s|} = \sup_{1 \leq k \leq K} \frac{|y_{k(j+1)} - y_{k(j)}|}{|t_{j+1} - t_j|} \leq \frac{\eta}{\tau}$$

but that it does not necessarily belong to $\mathcal{B}_R^\alpha([0,T],\mathbb{R})$.

The number of such functions f_k is bounded by the number of J-uples $(y_{k(j)})_{1 \leq j \leq J}$ such that $|y_{k(j+1)} - y_{k(j)}| \leq \eta$ for $1 \leq j \leq J-1$, that is the number of J-uples $(k(j))_{1 \leq j \leq J}$ such that $|k(j+1) - k(j)| \leq 1$ for $1 \leq j \leq J-1$. Such J-uples are obtained by choosing $k(1)$ among $2K$ values, then $k(2)$ among 3 values for $-K+2 \leq k(1) \leq +K-1$ or 2 values for $k(1) = -K+1$ and $+K$, and so on. Hence there exist at most $2\,K\,3^{J-1}$ such functions f_k.

If we now let K be the smallest integer larger or equal to $4\,\frac{R}{r}$ and J such that $KT^\alpha < J^\alpha$, then

$$\|f - f_k\|_\infty \leq \frac{r}{2}.$$

Indeed, given t in $[0,T]$, there exists some integer number j in $[1,J]$ such that t belong to $[t_j - \frac{\tau}{2}\,, t_j + \frac{\tau}{2}]$, so that

$$\begin{aligned} |f(t) - f_k(t)| &\leq |f(t) - f(t_j)| + |f(t_j) - f_k(t_j)| + |f_k(t_j) - f_k(t)| \\ &\leq R\,|t - t_j|^\alpha + |f(t_j) - y_{j(k)}| + \frac{\eta}{\tau}\,|t_j - t| \leq R\Big(\frac{\tau}{2}\Big)^\alpha + \frac{\eta}{2} + \frac{\eta}{\tau}\frac{\tau}{2} \leq 2\eta \leq \frac{r}{2}. \end{aligned}$$

Hence we can cover $\mathcal{B}_R^\alpha([0,T],\mathbb{R})$ by less than $2\,K\,3^{J-1}$ balls of radius $\dfrac{r}{2}$ of the metric space $\mathcal{C}([0,T],\mathbb{R})$ equipped with the uniform norm, and if we let J and K be the smallest integers larger or equal to $5^{\frac{1}{\alpha}}\,T\,(\dfrac{R}{r})^{\frac{1}{\alpha}}$ and $4\,\dfrac{R}{r}$ respectively, then $KT^\alpha < J^\alpha$ holds true, and

$$2\,K\,3^{J-1} \leq 10\,\frac{R}{r}\,3^{5^{\frac{1}{\alpha}}\,T\,(\frac{R}{r})^{\frac{1}{\alpha}}}.$$

1. 2. From this we now deduce the upper bound in the general case $d \geq 1$.

Let F be a given function in $\mathcal{B}_R^\alpha([0,T],\mathbb{R}^d)$ with components $F_i \in \mathcal{B}_R^\alpha([0,T],\mathbb{R})$, $1 \leq i \leq d$. Let now J and K be the smallest integers larger or equal to $5^{\frac{1}{\alpha}}\,T\,(\sqrt{d}\,\dfrac{R}{r})^{\frac{1}{\alpha}}$ and $4\,\sqrt{d}\,\dfrac{R}{r}$ respectively. With each i, we associate an integer k_i in $[1, 2\,K\,3^{J-1}]$ such that

$$\|F_i - f_{k_i}\|_\infty \leq \frac{r}{2\,\sqrt{d}}$$

where the f_k are the functions in $\mathcal{C}([0,T],\mathbb{R})$ defined in the first step (relatively to $\dfrac{r}{\sqrt{d}}$ instead of r).

Then the function $F_{k_1,\cdots,k_d}$ with components f_{k_i} for $1 \leq i \leq d$ belongs to $\mathcal{C}([0,T],\mathbb{R}^d)$ and satisfies $\|F - F_{k_1,\cdots,k_d}\|_\infty \leq r/2$. Moreover there are at most $(2\,K\,3^{J-1})^d$ such functions $F_{k_1,\cdots,k_d}$.

Consequently we can cover $\mathcal{B}_R^\alpha([0,R],\mathbb{R}^d)$ by less than $(2\,K\,3^{J-1})^d$ balls of radius $\dfrac{r}{2}$ of the metric space $\mathcal{C}([0,T],\mathbb{R}^d)$ equipped with the uniform norm, whence by less than $\big(2\,K\,3^{J-1}\big)^d$ balls of radius r of the metric space $\mathcal{B}_R^\alpha([0,T],\mathbb{R}^d)$ equipped with the uniform norm.

This concludes the proof of the upper bound of the covering number $\mathcal{N}(\mathcal{B}_R^\alpha([0,T],\mathbb{R}^d), r)$.

2. We now turn to the lower bound.

2.1. We first consider the case $d = 1$.

We can give different types of lower bounds by considering special functions of the type f_k defined in the first step. Here, for instance, we give the detail for one of them.

Given some non-zero integer J, we let $\tau = \dfrac{T}{J}$ and $\eta = \tau^\alpha\,R$, and then

$$\begin{array}{rclcrl} t_j & = & (j-\frac{1}{2})\,\tau\,, & j \in \mathbb{N}\,, & 1 & \leq j \leq J, \\ y_k & = & (k-\frac{1}{2})\,\eta\,, & k \in \mathbb{N}\,, & -\tau^{-\alpha}+\frac{1}{2} & \leq k \leq \tau^{-\alpha}+\frac{1}{2}. \end{array}$$

From a map $k : [1,J]\cap\mathbb{N} \to [0,1]\cap\mathbb{N}$, we define as above the function $f_k : [0,T] \to [y_0, y_1]$ affine on every interval of the subdivision $(0, t_1, \cdots, t_J, T)$ and such that

$$\begin{array}{rcll} f_k(0) & = & f_k(t_1) & \\ f_k(t_j) & = & y_{k(j)}, & 1 \leq j \leq J \\ f_k(T) & = & f_k(t_J). & \end{array}$$

Given some integer l such that $-\tau^{-\alpha}+\dfrac{1}{2}\leq l\leq \tau^{-\alpha}-\dfrac{1}{2}$, we define the function $f_{kl}:[0,T]\to[y_l,y_{l+1}]$ such that

$$f_{kl}(t)=f_k(t)+l\,\eta.$$

Then f_{kl} belongs to $\mathcal{B}^\alpha_R([0,T],\mathbb{R})$ and $\|f_{kl}-f_{k'l'}\|_\infty\geq\eta$ if $f_{kl}\neq f_{k'l'}$.

If for instance $r<\inf(R,2^{-1}T^\alpha R)$ and $J+1$ is the smallest integer larger or equal to $2^{-\frac{1}{\alpha}}T\,(\dfrac{R}{r})^{\frac{1}{\alpha}}$, then $\|f_{kl}-f_{k'l'}\|_\infty>2r$ if $f_{kl}\neq f_{k'l'}$.

Thus we have found $L\,2^J$ elements in $\mathcal{B}^\alpha_R([0,T],\mathbb{R})$ mutually distant of at least $2r$ in uniform norm, where L is the number of integers l such that $-\tau^{-\alpha}+\dfrac{1}{2}\leq l\leq\tau^{-\alpha}-\dfrac{1}{2}\cdot$ Thus

$$\mathcal{N}(\mathcal{B}^\alpha_R([0,T],\mathbb{R}),r)\ \geq\ L\,2^J.$$

But

$$L\ >\ 2\big((\tau^{-\alpha}-\frac{1}{2})-1\big)+1\ =\ 2\tau^{-\alpha}-2\ \geq\ \big((\frac{R}{r})^{\frac{1}{\alpha}}-\frac{2^{\frac{1}{\alpha}}}{T}\big)^\alpha-2\ \geq\ \frac{R}{r}-\frac{2}{T^\alpha}-2.$$

If moreover, for instance, $r\leq\dfrac{T^\alpha}{4\,T^\alpha+4}R$, then $L\geq\dfrac{R}{2r}$ and

$$\mathcal{N}(\mathcal{B}^\alpha_R([0,T],\mathbb{R}),r)\ \geq\ \frac{1}{4}\,\frac{R}{r}\,2^{2^{-\frac{1}{\alpha}}T\,(\frac{R}{r})^{\frac{1}{\alpha}}}.$$

2. 2. From this we now deduce the lower bound in the general case $d\geq1$.

The $L^d\,2^{dJ}$ functions $F_{k_1l_1,\cdots,k_dl_d}$ with components $f_{k_jl_j}$ for $j=1,\cdots,d$ where $f_{k_jl_j}$ have been defined in the first step, belong to $\mathcal{B}^\alpha_R([0,T],\mathbb{R}^d)$ and are mutually distant of at least $2\sqrt{d}\,r$.

This concludes the argument for the lower bound of the number $\mathcal{N}(\mathcal{B}^\alpha_R([0,T],\mathbb{R}^d),r)$. $\square$

We now turn to the covering number of the corresponding space of probability measures : given a Polish metric space (E,d), $p\geq1$ and $\delta>0$, we denote $\mathcal{N}_p(\mathcal{P}(E),\delta)$ the covering number of $\mathcal{P}(E)$ for the W_p distance.

Then we have the following general result which is proven in [22] (see also [43], [66]) :

Theorem 7.9. *Let (E,d) be a Polish metric space with finite diameter D, p and δ some real numbers with $p\geq1$ and $0<\delta<D$. Then the covering number $\mathcal{N}_p(\mathcal{P}(E),\delta)$ of $\mathcal{P}(E)$ satisfies*

$$\mathcal{N}_p(\mathcal{P}(E),\delta)\leq\big(8\,e\,\frac{D}{\delta}\big)^{p\mathcal{N}(E,\frac{\delta}{2})}$$

where $\mathcal{N}(E,\delta)$ is the covering number of E.

Note that if $\delta\geq D$, we simply have $\mathcal{N}_p(\mathcal{P}(E),\delta)=1$ since the Wasserstein distance between any two probability measures on E is at most D.

Since $\mathcal{B}^\alpha_R$ equipped with the metric defined by the uniform norm is a Polish metric space with finite diameter $2R$, we deduce the following result :

Theorem 7.10. *Let $d \geq 1$, p, T, R, δ and α be some positive numbers with $p \geq 1$, $\delta < 2R$ and $\alpha \leq 1$. Let also $\mathcal{B}_R^\alpha = \{f \in \mathcal{C}^\alpha ; \|f\|_\alpha \leq R\}$ be equipped with the uniform norm. Then the space $\mathcal{P}(\mathcal{B}_R^\alpha)$ of probability measures on $\mathcal{B}_R^\alpha$ can be covered by $\mathcal{N}_p(\mathcal{P}(\mathcal{B}_R^\alpha), \delta)$ balls of radius δ in Wasserstein distance W_p, with*

$$\mathcal{N}_p(\mathcal{P}(\mathcal{B}_R^\alpha), \delta) \leq \left(16\, e\, R\delta^{-1}\right)^{p\,(20\,\sqrt{d}\,R\,\delta^{-1})^d\, 3^{10^{\frac{1}{\alpha}}\, d^{1+\frac{1}{2\delta}}\, T\,(R\delta^{-1})^{\frac{1}{\alpha}}}}.$$

For $\delta \geq 2R$, we have

$$\mathcal{N}_p(\mathcal{P}(\mathcal{B}_R^\alpha), \delta) = 1.$$

Annexe

Espaces de mesures

Dans cette annexe nous obtenons des propriétés de séparabilité, métrisabilité et complétude d'espaces de mesures de probabilité à poids ; nous considérons en particulier le cas de tels espaces munis d'une distance de Wasserstein et montrons qu'ils sont complets et séparables s'il en est de même de l'espace sous-jacent.

Introduction

If (X, τ) is a topological space, i.e. a set X with a topology τ which will simply be denoted X if there is no ambiguity, and ω a real-valued continuous function on X, bounded by below by a positive constant, we denote $\mathcal{P}_\omega(X)$ the set of Borel probability measures μ on X such that

$$\int_X \omega(x)\, d\mu(x) \ < +\infty.$$

In the framework of measures we equip this set $\mathcal{P}_\omega(X)$ with the natural weak topology defined by the set $\mathcal{C}_{b\omega}(X)$ of real-valued continuous functions f on X such that $\omega^{-1} f$ be bounded on X ; this topology, that will be denoted w-$\mathcal{C}_{b\omega}(X)$, is defined by the seminorms

$$\mu \mapsto \sup_{i=1,\dots,n} |\int_X f_i(x)\, d\mu(x)|$$

for any finite family $f_1, \dots, f_n$ of functions of $\mathcal{C}_{b\omega}(X)$.

In this work we are concerned with some separability, metrizability and completeness properties of the topological space $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$.

In a first part we give the following general result :

Theorem A.1. *If the topological space (X, τ) is separable (resp. separable and metrizable, resp. separable, metrizable and topologically complete), then so is $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$.*

Conversely if $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$ is separable (resp. separable, metrizable and topologically complete), then so is (X, τ) if (X, τ) is a priori metrizable.

We recall that a topological space (X, τ) is separable if X has a countable dense subset, (X, τ) is metrizable if the topology τ is defined by a metric d on X, and (X, τ) is metrizable and topologically complete if it is metrized by a metric d such that the associated metric space (X, d) is complete. A separable, metrizable and topologically complete space is called a Polish space.

We consider here these sets of probability measures on such spaces, having in mind in particular some applications to problems arising in statistical physics, where, as in chapter 7, it is useful to handle probability measures on infinite-dimensional spaces such as the Wiener space of $\mathbb{R}^d$-valued continuous functions on some interval $[0, T]$, or some sets of probability measures on some phase space.

For $\omega = 1$, that is without weight, some results of this type are known in different ways : for instance [13], [47], [92], ... build some explicit distances on the set of Borel probability measures on a metric space, whereas [24], [42], [100], ... give a functional approach for Radon measures.

These results are respectively proven in sections A.2, A.3 and A.4 by functional methods. We first prove similar results for the set of finite nonnegative Borel measures on X equipped with the (narrow) weak topology defined by the set of bounded real-valued continuous functions on X, and then we deduce the announced properties by an homothetic transformation on measures and functions.

Then, in a second part, we shall more specifically consider the case when (X, d) is a metric space and ω is the weight

$$\omega \;=\; 1 \;+\; d(x_0, \cdot)^p$$

where p is a positive real number and x_0 belongs to X. The associated space $\mathcal{P}_\omega(X)$ is independent of x_0 and will be denoted $\mathcal{P}_p(X)$: then $\mathcal{P}_p(X)$ is the set of Borel probability measures μ on X with finite moment of order p (relatively to d), that is, such that

$$\int_X d(x_0, x)^p \, d\mu(x) \; < +\infty.$$

In the same way the space $\mathcal{C}_{b\omega}(X)$ will be denoted $\mathcal{C}_{bp}(X)$.

The $w\text{-}\mathcal{C}_{bp}(X)$ topology can be defined on the set $\mathcal{P}_p(X)$ by different metrics (see Remark A.11 and below). In section A.5 we shall see how this set $\mathcal{P}_p(X)$ is metrized by the Wasserstein distance W_p, which plays a basic role in mass transportation problems. If (X, d) is a separable complete metric space, the map W_p defined on $\mathcal{P}_p(X) \times \mathcal{P}_p(X)$ by

$$\begin{aligned}
W_p(\mu, \nu) \;&=\; \inf_\pi \Big(\iint_{X \times X} d(x, y)^p \, d\pi(x, y) \Big)^{\frac{1}{p}} \qquad && \text{if} \quad 1 \leq p \\
W_p(\mu, \nu) \;&=\; \inf_\pi \iint_{X \times X} d(x, y)^p \, d\pi(x, y) \qquad && \text{if} \quad 0 < p < 1,
\end{aligned}$$

where π runs over the set of probability measures on $X \times X$ with marginals μ and ν, is a metric on the set $\mathcal{P}_p(X)$ (see [3], [94] or [111] for instance), and in this case we shall prove

Theorem A.2. *If (X, d) is a separable complete metric space and p a positive number, then*
1 - the w-$\mathcal{C}_{bp}(X)$ topology is defined on the set $\mathcal{P}_p(X)$ by the metric W_p,
2 - the metric space $(\mathcal{P}_p(X), W_p)$ is separable and complete.

A.1 Notation

In this section we fix some notation and recall some classical and useful results on Borel measures.

Given a topological space (X, τ) which will in general be simply called X, we denote $\mathcal{M}_b^+(X)$ the set of finite nonnegative Borel measures on X, that is, the set of finite nonnegative measures on the σ-algebra generated by the topology τ, and $\mathcal{P}(X)$ the subset of $\mathcal{M}_b^+(X)$ of probability measures, that is, with total mass 1.

Denoting $\mathcal{C}_b(X)$ the set of bounded real-valued continuous functions on X, we equip the set $\mathcal{M}_b^+(X)$ with the topology defined by the seminorms

$$\mu \mapsto \sup_{i=1,\dots,n} \Big| \int_X f_i(x)\, d\mu(x) \Big|$$

for any finite family $f_1, \dots, f_n$ in $\mathcal{C}_b(X)$. This topology is denoted w-$\mathcal{C}_b(X)$ and often called the narrow topology on $\mathcal{M}_b^+(X)$.

Thus for any bounded real-valued continuous function f on X, the map

$$\mu \mapsto \int_X f(x)\, d\mu(x)$$

is a real-valued continuous function on $(\mathcal{M}_b^+(X), w\text{-}\mathcal{C}_b(X))$. More generally, if f is only lower semicontinuous, we recall the classical result :

Proposition A.3. *If (X, τ) is metrizable, then for any nonnegative real-valued lower semicontinuous function f on X, the map*

$$\mu \mapsto \int_X f(x)\, d\mu(x)$$

is a $\mathbb{R} \cup \{+\infty\}$-valued lower semicontinuous function on $(\mathcal{M}_b^+(X), w\text{-}\mathcal{C}_b(X))$.

Proof. Assuming that the topology τ is defined by a metric d on X, we can approach f by the nondecreasing sequence of continuous bounded functions $(f_n)_n$ defined on X by

$$f_n(x) = \inf\Big\{ \inf_{y \in X} [f(y) + n\, d(x, y)], n \Big\}.$$

Then for any $\mu \in \mathcal{M}_b^+(X)$, the monotone convergence theorem ensures

$$\int_X f(x)\, d\mu(x) = \sup_n \int_X f_n(x)\, d\mu(x).$$

Since moreover for any integer n the map

$$\mu \mapsto \int_X f_n(x)\, d\mu(x)$$

is a real-valued continuous function on $(\mathcal{M}_b^+(X), w\text{-}\mathcal{C}_b(X))$, this concludes the argument. □

In particular, if (X,τ) is metrizable and $(\mu_n)_n$ converges to μ in $(\mathcal{M}_b^+(X), w\text{-}\mathcal{C}_b(X))$, then for any nonnegative real-valued lower semicontinuous function f on X

$$\int_X f(x)\, d\mu(x) \leq \liminf_{n\to+\infty} \int_X f(x)\, d\mu_n(x).$$

The subset $\mathcal{P}(X)$ will be endowed with the induced w-$\mathcal{C}_b(X)$ topology.

In particular, considering the subset $\mathcal{D}(X) = \{\delta_x;\, x \in X\}$ of $\mathcal{P}(X)$ where, given $x \in X$, the point mass δ_x is defined by $\delta_x(B) = 1$ if $x \in B$ and 0 otherwise, for all Borel sets B of X, we recall the following result (see [13] for instance) :

Proposition A.4. *If (X,τ) is metrizable, then the topological space $(\mathcal{D}(X), w\text{-}\mathcal{C}_b(X))$ is homeomorphic to (X,τ) and the subset $\mathcal{D}(X)$ is sequentially closed in $(\mathcal{P}(X), w\text{-}\mathcal{C}_b(X))$.*

Proof. For the first point the map $x \mapsto \delta_x$ can be proven to be an homeomorphism from (X,τ) onto $(\mathcal{D}(X), w\text{-}\mathcal{C}_b(X))$.

Then let $(x_n)_n$ be a sequence of points in X such that the sequence $(\delta_{x_n})_n$ converges to some μ in $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$. Then there exists some subsequence of $(x_n)_n$ which converges to some x in (X,τ), so that $\mu = \delta_x$ by the first point.

Indeed let us assume that $(x_n)_n$ does not have any convergent subsequence. Then any subset F of the subsequence $(x_{n'})_{n'}$ is a closed subset of (X,τ). The characteristic function $\mathbf{1}_{X\setminus F}$ of the open set $X \setminus F$ is lower semicontinuous, so by Proposition A.3

$$1 - \mu(F) = \int_X \mathbf{1}_{X\setminus F}(x)\, d\mu(x) \leq \liminf_{n'} \int_X \mathbf{1}_{X\setminus F}(x)\, d\delta_{x'_n}(x) = 0.$$

Consequently $\mu(F) = 1$ for every infinite subsequence $(x_{n'})_{n'}$, which is impossible for such a probability measure μ. This concludes the argument. □

We now recall an important result about the sequential compactness of subsets of the space $(\mathcal{P}(X), w\text{-}\mathcal{C}_b(X))$. This property is linked to the following notion of tightness which proves important both in the theory of w-$\mathcal{C}_b(X)$ convergence and its applications.

Definition A.5. Let (X,τ) be a topological space. A subset $\mathcal{F}$ of $\mathcal{P}(X)$ is uniformly tight if for every $\varepsilon > 0$ there exists a compact set K in X such that $\mu(K) > 1 - \varepsilon$ for all $\mu \in \mathcal{F}$.

Then we have the following result (see [13] for instance) :

Proposition A.6 (Prokhorov). *Let (X,τ) be a metrizable space. Then any uniformly tight subset of $\mathcal{P}(X)$ is relatively sequentially compact for the w-$\mathcal{C}_b(X)$ topology.*

Prokhorov also proved a converse of this result under suitable extra conditions : if (X,τ) is a Polish space, then any relatively sequentially compact subset of $\mathcal{P}(X)$ for the w-$\mathcal{C}_b(X)$ topology is uniformly tight. As a particular case we have the following result, which can be proven directly :

Proposition A.7 (Ulam). *Let (X,τ) be a Polish space. Then any finite family of $\mathcal{P}(X)$ is uniformly tight.*

In the sequel we shall consider some weighted Borel measures. More precisely we let ω be a real-valued continuous function on X, bounded by below by a positive constant, called a weight. Then we denote $\mathcal{M}_{b\omega}^+(X)$ the set of finite nonnegative Borel measures μ on X such that $\omega\,\mu \in \mathcal{M}_b^+(X)$ and $\mathcal{P}_\omega(X)$ the set of probability measures on X that belong to $\mathcal{M}_{b\omega}^+(X)$.

Denoting $\mathcal{C}_{b\omega}(X)$ the set of functions f on X such that $\omega^{-1} f \in \mathcal{C}_b(X)$, we equip the set $\mathcal{M}_{b\omega}^+(X)$ with the topology defined by the seminorms

$$\mu \mapsto \sup_{i=1,\dots,n} \left|\int_X f_i(x)\, d\mu(x)\right|$$

for any finite family $f_1, \dots, f_n$ in $\mathcal{C}_{b\omega}(X)$. This topology is denoted w-$\mathcal{C}_{b\omega}(X)$. The subset $\mathcal{P}_\omega(X)$ will be endowed with the induced w-$\mathcal{C}_{b\omega}(X)$ topology.

A.2 Separability

In this section we prove the first statement in Theorem A.1, that is, the topological space $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$ is separable when so is (X,τ), and conversely if (X,τ) is a priori metrizable.

First of all we prove

Proposition A.8. *If (X,τ) is a separable topological space, then so is $(\mathcal{M}_b^+(X), w\text{-}\mathcal{C}_b(X))$.*

Proof. If (x_n) is a sequence dense in (X,τ), then the countable family of combinations of point masses δ_{x_n} at x_n with nonnegative rational coefficients is dense in $(\mathcal{M}_b^+(X), w\text{-}\mathcal{C}_b(X))$.

Let indeed $\mu \in \mathcal{M}_b^+(X)$ with $\mu \neq 0$, $\varepsilon > 0$ and $f_1, \dots, f_n \in \mathcal{C}_b(X)$ such that $0 < f_j < 1$ (as we may assume).

Given an integer $k \geq \dfrac{2\,+\,2\,\mu(X)}{\varepsilon}$, let

$$X_{ji} = \left\{x \in X; \frac{i-1}{k} \leq f_j(x) < \frac{i}{k}\right\}$$

for $j = 1, \ldots, n$ and $i = 1, \ldots, k$. Since $0 < f_j < 1$, we have

$$\sum_{i=1}^{k} \frac{i-1}{k} \mu(X_{ji}) \leq \int_X f_j(x)\, d\mu(x) < \sum_{i=1}^{k} \frac{i}{k} \mu(X_{ji}).$$

Let then $(X_l)_{l \in L}$ a finite partition of X generated by these $(X_{ji})_{j,i}$ such that

$$X_{ji} = \bigcup_{l \in L_{ji}} X_l$$

for any j, i, and for some subsets L_{ji} of L such that $\bigcup_{i=1}^{k} L_{ji} = L$ for any j.

Let also $(q_l)_{l \in L}$ be some nonnegative rational numbers such that

$$|\mu(X_l) - q_l| < \frac{1}{k \operatorname{card} L}$$

for any l.

Then choosing any $y_l \in X_l$ for each $l \in L$, it can be checked that

$$|\sum_{i=1}^{k} \frac{i-1}{k} \mu(X_{ji}) - \sum_{l \in L} q_l f_j(y_l)| < \frac{1 + \mu(X)}{k}$$

for any $j = 1, \ldots, n$. But by density of the family $(x_n)_n$ and continuity of f_j at y_l, for any $\alpha > 0$ and $l \in L$ there exists some x_{n_l} such that

$$|f_j(y_l) - f_j(x_{n_l})| \leq \alpha$$

for any $j = 1, \ldots, n$. Then

$$\begin{aligned}
|\sum_{i=1}^{k} \frac{i-1}{k} \mu(X_{ji}) - \sum_{l \in L} q_l f_j(x_{n_l})| \;&\leq\; |\sum_{i=1}^{k} \frac{i-1}{k} \mu(X_{ji}) - \sum_{l \in L} q_l f_j(y_l)| + \sum_{l \in L} q_l |f_j(y_l) - f_j(x_{n_l})| \\
&< \frac{1 + \mu(X)}{k} + \left[\mu(X) + \frac{1}{k}\right] \alpha
\end{aligned}$$

which is bounded by above by $\dfrac{2 + \mu(X)}{k}$ provided $\alpha \leq \dfrac{1}{1 + k\,\mu(X)}$.

Consequently the measure

$$m = \sum_{l \in L} q_l \delta_{x_{n_l}}$$

satisfies

$$\left| \int_X f_j(x)\, d\mu(x) - \int_X f_j(x)\, dm(x) \right| < \frac{2 + 2\,\mu(X)}{k}$$

for $j = 1, \ldots, n$, that is, m belongs to the neighbourhood

$$\left\{\nu \in \mathcal{M}_b^+(X);\ \left|\int_X f_j(x)\, d\mu(x) - \int_X f_j(x)\, d\nu(x)\right| < \varepsilon,\ j = 1, \ldots, n\right\}$$

of μ since we have chosen $k \geq \dfrac{2\,+\,2\mu(X)}{\varepsilon}$. □

Then the following result can be proven from the previous proposition :

Proposition A.9. *If (X, τ) is a separable topological space and ω a weight as above, then $(\mathcal{M}_{b\omega}^+(X), w\text{-}\mathcal{C}_{b\omega}(X))$ also is separable.*

Proof. It is due to the fact that the map

$$h\,:\,\mu \mapsto \omega^{-1}\,\mu$$

is an homeomorphism from $(\mathcal{M}_b^+(X), w\text{-}\mathcal{C}_b(X))$ onto $(\mathcal{M}_{b\omega}^+(X), w\text{-}\mathcal{C}_{b\omega}(X))$. □

Hence, if (X, τ) is a separable topological space, then so is the subspace $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$, which concludes the proof of the sufficient condition of Theorem A.1 in this first case.

Conversely, if $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$ is a separable topological space, then so is its subspace $(\mathcal{D}(X), w\text{-}\mathcal{C}_{b\omega}(X))$ and then $(\mathcal{D}(X), w\text{-}\mathcal{C}_b(X))$. Consequently, if the topological space (X, τ) is metrizable, the first result of Proposition A.4 ensures the necessary condition in Theorem A.1 in this first case.

A.3 Metrizability

In this section we prove that the topological space $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$ is metrizable and separable when so is (X, τ).

First at all we prove

Proposition A.10. *If (X, τ) is a separable metrizable space, then so is $(\mathcal{M}_b^+(X), w\text{-}\mathcal{C}_b(X))$.*

Proof. Following for instance the ideas developed in [42] or [100] in the framework of Radon measures, we give an abstract functional approach of this result in the following way.

Let d be a distance on X which metrizes the topology τ.

1 - We first assume that (X, d) is a compact metric space.

We denote $\mathcal{M}(X)$ (resp. $\mathcal{M}^+(X)(= \mathcal{M}_b^+(X))$) the set of signed (resp. nonnegative) Borel measures on X and $\mathcal{C}(X)\,(= \mathcal{C}_b(X))$ the set of real-valued continuous functions on X, equipped with the supremum norm

$$\|f\|_{\mathcal{C}(X)} = \sup_x |f(x)|.$$

Riesz representation theorem ensures that the map $\mu \mapsto I_\mu$, where I_μ is defined on $\mathcal{C}(X)$ by

$$I_\mu(f) = \int_X f(x)\, d\mu(x),$$

is an isometry from the set $\mathcal{M}(X)$ with norm $\|\mu\|_{\mathcal{M}(X)} = |\mu|(X)$ onto the topological dual of $(\mathcal{C}(X), \|\cdot\|_{\mathcal{C}(X)})$ with dual norm.

Since the space $(\mathcal{C}(X), \|\cdot\|_{\mathcal{C}(X)})$ is a separable Banach space, Banach-Alaoglu theorem ensures then that the unit ball

$$\mathcal{M}^1(X) = \{\mu \in \mathcal{M}(X);\ |\mu|(X) \leq 1\}$$

of the Banach space $(\mathcal{M}(X), \|\cdot\|_{\mathcal{M}(X)})$ is compact and metrizable for the w-$\mathcal{C}(X)$ topology.

The set

$$\mathcal{M}^{+1}(X) = \big\{\mu \in \mathcal{M}^+(X);\ \mu(X) \leq 1\big\}$$

is a closed subset of $(\mathcal{M}^1(X), w\text{-}\mathcal{C}(X))$. Thus $(\mathcal{M}^{+1}(X), w\text{-}\mathcal{C}(X))$ is also compact and metrizable, hence Polish.

Then the set

$$\overset{\circ}{\mathcal{M}}{}^{+1}(X) = \big\{\mu \in \mathcal{M}^+(X);\ \mu(X) < 1\big\}$$

is an open subset of $(\mathcal{M}^{+1}(X), w\text{-}\mathcal{C}(X))$ since for any $f \in \mathcal{C}(X)$, in particular $f = 1$, the map

$$\mu \mapsto \int_X f(x)\, d\mu(x)$$

is continuous from $(\mathcal{M}^{+1}(X), w\text{-}\mathcal{C}(X))$ into $(\mathbb{R}, |\cdot|)$. Thus $(\overset{\circ}{\mathcal{M}}{}^{+1}(X), w\text{-}\mathcal{C}(X))$ also is Polish.

Finally, as the map

$$\mu \mapsto \frac{\mu}{1 + \mu(X)}$$

is an homeomorphism from $(\mathcal{M}^+(X), w\text{-}\mathcal{C}(X))$ onto $(\overset{\circ}{\mathcal{M}}{}^{+1}(X), w\text{-}\mathcal{C}(X))$, it follows that $(\mathcal{M}^+(X), w\text{-}\mathcal{C}(X))$ is a Polish space too.

2 - We now assume that (X, d) is a separable metric space.

Possibly replacing the distance d by a distance e defining the same topology on X, we may consider X as a subset of a compact metric space $(\tilde{X}, \tilde{e})$ (the completion of (X, e)) such that the identity map i from X into $\tilde{X}$ be an homeomorphism from (X, e) onto $(Y, \tilde{e})$ where Y is a dense subset of $\tilde{X}$. In particular the Borel sets of X are the $i^{-1}(\tilde{B}) = \tilde{B} \cap X$ where $\tilde{B}$ is a Borel set of $\tilde{X}$.

Let us consider the map

$$I : \mu \mapsto \mu \circ i^{-1}$$

from $M_b^+(X)$ to $M^+(\tilde{X})$, where $\mu \circ i^{-1}$ is the image measure of μ by i defined by

$$(\mu \circ i^{-1})(\tilde{B}) = \mu(i^{-1}(\tilde{B})) \ \big(= \mu(\tilde{B} \cap X)\big)$$

for any Borel set $\tilde{B}$ of $\tilde{X}$, or by

$$\int_{\tilde{X}} \tilde{f}(\tilde{x})\, d(\mu \circ i^{-1})(\tilde{x}) = \int_X (\tilde{f} \circ i)(x)\, d\mu(x) \left(= \int_X \tilde{f}\,|_X\,(x)\, d\mu(x)\right)$$

for any measurable function $\tilde{f}$ from $\tilde{X}$ into $\mathbb{R}$.

First of all, the map I is continuous from $(M_b^+(X), w\text{-}\mathcal{C}_b(X))$ into $(M^+(\tilde{X}), w\text{-}\mathcal{C}(\tilde{X}))$ since, by the above formula, for any $\mu \in M_b^+(X)$ and $\tilde{f} \in C(\tilde{X})$,

$$\int_{\tilde{X}} \tilde{f}(\tilde{x})\, d(I\mu)(\tilde{x}) = \int_X f(x)\, d\mu(x)$$

where f is the function of $C_b(X)$ defined by $f = \tilde{f}\,|_X$.

Then this map is one-to-one. Let indeed μ and $\nu \in M_b^+(X)$ such that $I\mu = I\nu$. Then, for any Borel set B of X, there exists some Borel set $\tilde{B}$ of $\tilde{X}$ such that $B = \tilde{B} \cap X$; then

$$\mu(B) = \mu(\tilde{B} \cap X) = (I\mu)(\tilde{B}) = (I\nu)(\tilde{B}) = \nu(\tilde{B} \cap X) = \nu(B),$$

so that $\mu = \nu$.

Finally, letting $M^+(\tilde{X}, X) = I(M_b^+(X))$, I^{-1} is continuous from $(M^+(\tilde{X}, X), w\text{-}\mathcal{C}(\tilde{X}))$ into $(M_b^+(X), w\text{-}\mathcal{C}_b(X))$. To prove this point, as $(M^+(\tilde{X}, X), w\text{-}\mathcal{C}(\tilde{X}))$ is metrizable (and separable) because so is $(M^+(\tilde{X}), w\text{-}\mathcal{C}(\tilde{X}))$ by step 1, it is sufficient to prove that if $(\mu_n)_n$ and μ in $M_b^+(X)$ are such that $I\mu = \lim\limits_{n\to+\infty} I\mu_n$ in $(M^+(\tilde{X}), w\text{-}\mathcal{C}(\tilde{X}))$, then $\displaystyle\int_X f(x)\, d\mu(x) = \lim_{n\to+\infty} \int_X f(x)\, d\mu_n(x)$ for all $f \in C_b(X)$.

Assuming for instance that $0 \leq f \leq 1$, we let f_0 and f_1 be the functions defined on $\tilde{X}$ by extending f outside X by 0 and 1 respectively, and we let $\overline{f_0}$ and $\underline{f_1}$ be the upper and lower semicontinuous functions on $\tilde{X}$ defined by

$$\overline{f_0}(x) = \limsup_{y\to x} f_0(y) \qquad \underline{f_1}(x) = \liminf_{y\to x} f_1(y).$$

Since $\underline{f_1}$ is lower semicontinuous on $\tilde{X}$, Proposition A.3 ensures that

$$\int_{\tilde{X}} \underline{f_1}(\tilde{x})\, d(I\mu)(\tilde{x}) \leq \liminf_{n\to+\infty} \int_{\tilde{X}} \underline{f_1}(\tilde{x})\, d(I\mu_n)(\tilde{x}).$$

In the same way

$$\int_{\tilde{X}} \overline{f_0}(\tilde{x})\, d(I\mu)(\tilde{x}) \geq \limsup_{n\to+\infty} \int_{\tilde{X}} \overline{f_0}(\tilde{x})\, d(I\mu_n)(\tilde{x}).$$

But $\overline{f_0} = \underline{f_1} = f$ on X, so

$$\limsup_{n\to+\infty} \int_X f(x)\, d\mu_n(x) \leq \int_X f(x)\, d\mu(x) \leq \liminf_{n\to+\infty} \int_X f(x)\, d\mu_n(x),$$

whence

$$\int_X f(x)\,d\mu(x) = \lim_{n\to+\infty} \int_X f(x)\,d\mu_n(x).$$

Hence the map I is an homeomorphism from $(M_b^+(X), w\text{-}\mathcal{C}_b(X))$ onto $(M^+(\tilde{X}, X), w\text{-}\mathcal{C}(\tilde{X}))$.

Consequently $(\mathcal{M}_b^+(X), w\text{-}\mathcal{C}_b(X))$ is metrizable and separable, with the distance associated with the distance on $\mathcal{M}^+(\tilde{X}, X)$. □

Remark A.11. This proposition can also be proven by explicitly building a distance on $\mathcal{M}_b^+(X)$.

If (X, τ) is a separable metrizable space, some distances can indeed be explicitly built on $\mathcal{P}(X)$, that define this $w\text{-}\mathcal{C}_b(X)$ topology.

For instance if the topology τ is defined by a metric d on X, the map ρ defined on $\mathcal{P}(X) \times \mathcal{P}(X)$ by

$$\rho(\mu, \nu) = \inf\{\varepsilon > 0;\ \nu(B) \leq \mu(B^\varepsilon) + \varepsilon \text{ for all Borel set } B\}$$

where

$$B^\varepsilon = \{x \in X;\ d(x, B) < \varepsilon\}$$

is a metric on $\mathcal{P}(X)$, called the Prokhorov metric, or sometimes the Lévy-Prokhorov metric. This distance is to be linked to the following equivalent definition of the $w\text{-}\mathcal{C}_b(X)$ topology on $\mathcal{P}(X)$: a fundamental set of neighbourhoods of μ for this topology is made up of the

$$\{\nu \in \mathcal{P}(X);\ \nu(F_i) \leq \mu(F_i) + \varepsilon,\ i = 1, \ldots, n\}$$

for a finite family of closed sets $F_1, \ldots, F_n$ of X and ε a positive number.

In the same way the map ρ defined on $\mathcal{P}(X) \times \mathcal{P}(X)$ by

$$\rho(\mu, \nu) = \sup\left\{\left|\int_X f(x)\,d\mu(x) - \int_X f(x)\,d\nu(x)\right|;\ f \in \mathcal{C}_b(X),\ \|f\|_{lip} \leq 1\right\},$$

where $\|.\|_{lip}$ is the Lipschitz norm defined by

$$\|f\|_{lip} = \max\Big(\sup_x |f(x)|, \sup_{x\neq y} \frac{|f(x) - f(y)|}{d(x, y)}\Big),$$

is also a metric on $\mathcal{P}(X)$, called the dual-bounded Lipschitz metric and often denotes d_{BL}.

If (X, τ) is a separable metrizable space, the $w\text{-}\mathcal{C}_b(X)$ topology on the set $\mathcal{P}(X)$ can then be metrized by these distances ρ (see [13], [42], [47], [92] for instance).

Then the map defined on $\mathcal{M}_b^+(X) \times \mathcal{M}_b^+(X)$ by

$$\begin{aligned}
\rho_1(\mu, \nu) &= \rho\big(\frac{\mu}{\mu(X)}, \frac{\nu}{\nu(X)}\big) + |\mu(X) - \nu(X)| \quad \text{if } \mu \text{ and } \nu \neq 0,\\
\rho_1(0, \nu) &= \nu(X) \quad \text{if } \mu = 0,
\end{aligned}$$

is a distance on the set $\mathcal{M}_b^+(X)$ which metrizes the $w\text{-}\mathcal{C}_b(X)$ topology.

Then the following result can be proven from the previous proposition :

Proposition A.12. *If (X, τ) is a separable metrizable space and ω a weight as above, then $(\mathcal{M}^+_{b\omega}(X), w\text{-}\mathcal{C}_{b\omega}(X))$ also is metrizable and separable.*

Proof. We only need to prove that $(\mathcal{M}^+_{b\omega}(X), w\text{-}\mathcal{C}_{b\omega}(X))$ is metrizable.

On one hand, as in the proof of Proposition A.9, we consider the map

$$h \,:\, \mu \mapsto \omega^{-1}\,\mu$$

from $\mathcal{M}^+_b(X)$ onto $\mathcal{M}^+_{b\omega}(X)$, which is an homeomorphism from $(\mathcal{M}^+_b(X), w\text{-}\mathcal{C}_b(X))$ onto $(\mathcal{M}^+_{b\omega}(X), w\text{-}\mathcal{C}_{b\omega}(X))$.

On the other hand by Proposition A.10, there exists a metric ρ_1 on $\mathcal{M}^+_b(X)$ such that the identity map i on $\mathcal{M}^+_b(X)$ be an homeomorphism from $(\mathcal{M}^+_b(X), w\text{-}\mathcal{C}_b(X))$ onto $(\mathcal{M}^+_b(X), \rho_1)$. We then define a metric ρ_ω on $\mathcal{M}^+_{b\omega}(X)$ by

$$\rho_\omega(\mu, \nu) \;=\; \rho_0(h^{-1}(\mu), h^{-1}(\nu)).$$

Note then that the map h is an isometric homeomorphism from $(\mathcal{M}^+_b(X), \rho_1)$ onto $(\mathcal{M}^+_{b\omega}(X), \rho_\omega)$ by definition of ρ_ω.

The identity map j on $\mathcal{M}^+_{b\omega}(X)$ is then an homeomorphism from $(\mathcal{M}^+_{b\omega}(X), w\text{-}\mathcal{C}_{b\omega}(X))$ onto $(\mathcal{M}^+_{b\omega}(X), \rho_\omega)$ since we have the following commutative diagram

$$\begin{array}{ccc}
(\mathcal{M}^+_b,\, w\text{-}\mathcal{C}_b(X)) & \stackrel{h}{\rightarrow} & (\mathcal{M}^+_{b\omega},\, w\text{-}\mathcal{C}_{b\omega}(X)) \\
\downarrow i & & \downarrow j \\
(\mathcal{M}^+_b, \rho_1) & \stackrel{h}{\rightarrow} & (\mathcal{M}^+_{b\omega}, \rho_\omega).
\end{array}$$

In particular $(\mathcal{M}^+_{b\omega}(X), w\text{-}\mathcal{C}_{b\omega}(X))$ is metrizable. □

Consequently the subspace $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$ is also metrizable by the distance ρ_ω, and separable, which concludes the proof of the second point in Theorem A.1.

Remark A.13. The argument developed in these two proofs would also give the separability of these measures spaces without calling up the results obtained in section A.2.

A.4 Completeness

We now prove the last point in Theorem A.1 in the same way we proved above the second point. First of all we have

Proposition A.14. *If (X, τ) is a Polish space, then so is $(\mathcal{M}^+_b(X), w\text{-}\mathcal{C}_b(X))$.*

Proof. Following the ideas of [42] and [100] and using the notations introduced in the second step of the proof of Proposition A.10, X is a G_δ of its completion $(\tilde{X}, \tilde{e})$, that is, $X = \bigcap_n U_n$ where (U_n) is a sequence of open sets of $(\tilde{X}, \tilde{e})$.

In particular X is a Borel set of $\tilde{X}$ and the Borel sets of X are the subsets of X which are Borel sets in $\tilde{X}$. In this case we have

$$M^+(\tilde{X}, X) = \{\tilde{\mu} \in M^+(\tilde{X});\ \tilde{\mu}(\tilde{X} \setminus X) = 0\}.$$

Indeed, if $\mu \in M_b^+(X)$, then

$$(I\mu)(\tilde{X} \setminus X) = (I\mu)(\tilde{X}) - (I\mu)(X) = \mu(X) - \mu(X) = 0.$$

Conversely, if $\tilde{\mu} \in M^+(\tilde{X})$ is such that $\tilde{\mu}(\tilde{X} \setminus X) = 0$, let $\tilde{\mu}_X$ be the measure induced on X by $\tilde{\mu}$, and defined by

$$\tilde{\mu}_X(B) = \tilde{\mu}(B)$$

for any Borel set B of X. Then, for any Borel set $\tilde{B}$ of $\tilde{X}$,

$$(I\tilde{\mu}_X)(\tilde{B}) = \tilde{\mu}_X(\tilde{B} \cap X) = \tilde{\mu}(\tilde{B} \cap X) = \tilde{\mu}(\tilde{B} \setminus (\tilde{B} \cap (\tilde{X} \setminus X))) = \tilde{\mu}(\tilde{B}) - \tilde{\mu}(\tilde{B} \cap (\tilde{X} \setminus X)) = \tilde{\mu}(\tilde{B}),$$

whence $\tilde{\mu} = I\tilde{\mu}_X$.

Then we decompose

$$\mathcal{M}^+(\tilde{X}, X) = \bigcap_{n,p} S_{n,p}$$

where

$$S_{n,p} = \{\tilde{\mu} \in M^+(\tilde{X});\ \tilde{\mu}(U_n) > \tilde{\mu}(\tilde{X}) - \frac{1}{p}\}$$

for any integer n, $p \geq 1$.

The indicator function $\mathbf{1}_{U_n}$ of the open set U_n is a lower semicontinuous function on $\tilde{X}$, so the map

$$\tilde{\mu} \mapsto \int_{\tilde{X}} \mathbf{1}_{U_n}(x)\, d\tilde{\mu}(x) = \tilde{\mu}(U_n)$$

is lower semicontinuous on $(M^+(\tilde{X}), w\text{-}\mathcal{C}(\tilde{X}))$ by Proposition A.3. Thus each $S_{n,p}$ is an open set of $(M^+(\tilde{X}), w\text{-}\mathcal{C}(\tilde{X}))$ and $M^+(\tilde{X}, X)$ is a G_δ in this space.

But from step 1 of the proof of Proposition A.10, this space is Polish and consequently so is its subspace $(M^+(\tilde{X}, X), w\text{-}\mathcal{C}(\tilde{X}))$.

Finally by the homeomorphism from $(\mathcal{M}_b^+(X), w\text{-}\mathcal{C}_b(X))$ onto $(\mathcal{M}^+(\tilde{X}, X), w\text{-}\mathcal{C}(\tilde{X}))$ introduced in step 2 of the proof of Proposition A.10, we can conclude that $(\mathcal{M}_b^+(X), w\text{-}\mathcal{C}_b(X))$ is a Polish space too. □

From this result we can now prove

Theorem A.1.3. *If (X, τ) is a Polish space and ω a weight as above, then the topological space $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$ also is a Polish space.*

Proof. From the previous proposition there exists a distance ρ_1 on $\mathcal{M}_b^+(X)$ which defines on $\mathcal{M}_b^+(X)$ the w-$\mathcal{C}_b(X)$ topology, and for which $(\mathcal{M}_b^+(X), \rho_1)$ is a separable complete metric space.

Then with the notations of the proof of Proposition A.12, the metric ρ_ω associated with ρ_1 defines on $\mathcal{M}_{b\omega}^+(X)$ the w-$\mathcal{C}_{b\omega}(X)$ topology, and the metric space $(\mathcal{M}_{b\omega}^+(X), \rho_\omega)$ is separable and complete since the map h is an isometry from $(\mathcal{M}_b^+(X), \rho_0)$ onto $(\mathcal{M}_{b\omega}^+(X), \rho_\omega)$.

Finally $\mathcal{P}_\omega(X)$ is a closed subset of $(\mathcal{M}_{b\omega}^+(X), \rho_\omega)$ since

$$\mathcal{P}_\omega(X) \;=\; \{\mu \in \mathcal{M}_{b\omega}^+(X), \int_X 1\, d\mu(x) \;=\; 1\}$$

and $x \mapsto 1$ belongs to $\mathcal{C}_{b\omega}(X)$.

Thus $(\mathcal{P}_\omega(X), \rho_\omega)$ is a separable complete metric space. This concludes the proof of the sufficient condition of Theorem A.1 in this case. □

Conversely, if $(x_n)_n$ is a Cauchy sequence in (X, d), then so is $(\delta_{x_n})_n$ in $(\mathcal{D}(X), w\text{-}\mathcal{C}_b(X))$ by Proposition A.4, and then in $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$ since in particular the identity map from $(\mathcal{D}(X), w\text{-}\mathcal{C}_b(X))$ onto $(\mathcal{D}(X), w\text{-}\mathcal{C}_{b\omega}(X))$ is continuous. Hence, if $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$ is complete, then $(\delta_{x_n})_n$ converges to some μ in $(\mathcal{P}_\omega(X), w\text{-}\mathcal{C}_{b\omega}(X))$, hence in $(\mathcal{P}(X), w\text{-}\mathcal{C}_b(X))$. Then Proposition A.4 ensures that $\mu = \delta_x$ for some $x \in X$, and that $(x_n)_n$ converges to x in (X, d).

Remark A.15. If (X, d) is a separable complete metric space, the Prokhorov distance or the dual-bounded Lipschitz distance defining the w-$\mathcal{C}_b(X)$ topology on $\mathcal{P}(X)$ are such that $(\mathcal{P}(X), \rho)$ is complete. Using the notation of Remark A.11, the distance ρ_1 associated with ρ is such that $(\mathcal{M}_b^+(X), \rho_1)$ is complete.

A.5 Wasserstein distances

In this section we let (X, τ) be a Polish space, d a lower semicontinuous distance on X, p a positive real number, and we consider the weight

$$\omega \;=\; 1 \,+\, d(x_0, \cdot)^p$$

for some x_0 in X. Then we simply denote $\mathcal{P}_p(X)$ the x_0-independent set of probability measures μ such that

$$\int_X d(x_0, x)^p\, d\mu(x) < +\infty.$$

The map W_p defined on $\mathcal{P}_p(X) \times \mathcal{P}_p(X)$ by

$$\begin{aligned} W_p(\mu, \nu) \;&=\; \inf_\pi \Big(\iint_{X\times X} d(x,y)^p\, d\pi(x,y) \Big)^{\frac{1}{p}} \qquad &\text{if } \; 1 \le p \\ W_p(\mu, \nu) \;&=\; \inf_\pi \iint_{X\times X} d(x,y)^p\, d\pi(x,y) \qquad &\text{if } \; 0 < p < 1, \end{aligned}$$

where π runs over the set $\Pi(\mu,\nu)$ of probability measures on $X \times X$ with marginals μ and ν, is a distance on $\mathcal{P}_p(X)$ (see [111] for instance).

First of all for the w-$\mathcal{C}_b(X)$ topology we have the following general result :

Proposition A.16. *With the above notation, let (μ_n) a sequence in $\mathcal{P}_p(X)$ converging to μ in $(\mathcal{P}(X), w\text{-}\mathcal{C}_b(X))$. Then*

$$\int_X d(x_0,x)^p \, d\mu(x) \le \liminf_{n\to+\infty} \int_X d(x_0,x)^p \, d\mu_n(x).$$

If moreover $\mu \in \mathcal{P}_p(X)$, then for all $\nu \in \mathcal{P}_p(X)$

$$W_p(\mu,\nu) \le \liminf_{n\to+\infty} W_p(\mu_n,\nu).$$

Proof. The first statement stems from Proposition A.3.

For the second point we first note that for any ρ_1 and ρ_2 in $\mathcal{P}_p(X)$ there exists some probability measure π on $X\times X$ achieving the infimum defining $W_p(\rho_1,\rho_2)$. Indeed $\Pi(\rho_1,\rho_2)$ is uniformly tight, so any minimizing sequence $(\pi_n)_n$ has a cluster point π by Prokhorov theorem on $X \times X$, which in turn achieves the infimum since, by Proposition A.3 applied to $X \times X$ and $d(.,.)$,

$$\iint_{X\times X} d(x,y)^p \, d\pi(x,y) \le \liminf_{n\to+\infty} \iint_{X\times X} d(x,y)^p \, d\pi_n(x,y) = W_p(\rho_1,\rho_2)^{\max(1,p)}.$$

Then let (n') be a subsequence such that $W_p(\mu_{n'},\nu)$ converge to $\liminf\limits_{n\to+\infty} W_p(\mu_n,\nu)$ as n' goes to infinity.

For each n', there exists by the above remark a probability measure $\pi_{n'}$ on $X \times X$, with marginals μ'_n and ν, such that

$$\iint_{X\times X} d(x,y)^p \, d\pi_{n'}(x,y) = W_p^p(\mu_{n'},\nu). \tag{A.1}$$

The sequence $(\mu'_n)_{n'}$ converges weakly, so is uniformly tight, so for any $\delta > 0$ there exists a compact K such that $\nu(K) \ge 1-\delta$ and $\mu_{n'}(K) \ge 1-\delta$ for any n'. Since $\pi_{n'}$ has marginals $\mu_{n'}$ and ν, this implies

$$\pi_{n'}(K \times K) = \mu_{n'}(K)\,\nu(K) \ge (1-\delta)^2 \ge 1-2\delta,$$

which means that the family $(\pi_{n'})_{n'}$ is tight. Thus by Prokhorov theorem there exists a subsequence $(\pi_{n''})_{n''}$ and a probability π on $X \times X$ such that $(\pi_{n''})_{n''}$ converges weakly to π as n'' goes to infinity. In particular by Proposition A.3 and (A.1) :

$$\begin{aligned}\iint_{X\times X} d(x,y)^p \, d\pi(x,y) &\le \liminf_{n''\to+\infty} \iint_{X\times X} d(x,y)^p \, d\pi_{n''}(x,y) \\ &= \lim_{n''\to+\infty} W_p^p(\mu_{n''},\nu) = \liminf_{n\to+\infty} W_p^p(\mu_n,\nu). \end{aligned} \tag{A.2}$$

Then $\pi_{n''}$ has marginals $\mu_{n''}$ and ν, so at the limit π has marginals μ and ν. In particular

$$W_p^p(\mu,\nu) \leq \iint_{X\times X} d(x,y)^p \, d\pi(x,y)$$

and hence from (A.2)

$$W_p^p(\mu,\nu) \leq \liminf_{n\to+\infty} W_p^p(\mu_n,\nu).$$

This concludes the argument. □

From now on we more specifically consider a separable complete metric space (X,d) and the Wasserstein distance W_p defined by the distance d. We shall prove, as announced in Theorem A.2, that this distance turns $(\mathcal{P}_p(X), W_p)$ into a separable complete metric space.

We let $\mathcal{C}_{bp}(X)$ denote the set of functions f on X such that $\big(1+d(x_0,\cdot)^p\big)^{-1} f \in \mathcal{C}_b(X)$ and equip $\mathcal{P}_p(X)$ with the topology defined by the seminorms

$$\mu \mapsto \sup_{i=1,\dots,n} \Big| \int_X f_i(x)\, d\mu(x) \Big|$$

for any finite family $f_1, \dots, f_n$ in $\mathcal{C}_{bp}(X)$. This topology is denoted w-$\mathcal{C}_{bp}(X)$, and is linked with the topology defined by the distance W_p by the following result (see [111] for instance) :

Proposition A.17. *Let (X,d) be a separable complete metric space, p a positive number, $(\mu_n)_n$ and μ in $\mathcal{P}_p(X)$. Then the following statements are equivalent :*

1. *$(\mu_n)_n$ converges to μ in $(\mathcal{P}_p(X), W_p)$,*
2. *$(\mu_n)_n$ converges to μ in $(\mathcal{P}_p(X), w\text{-}\mathcal{C}_{bp}(X))$,*
3. *$(\mu_n)_n$ converges to μ in $(\mathcal{P}(X), w\text{-}\mathcal{C}_b(X))$ and $\Big(\int_X d(x_0,x)^p \, d\mu_n(x)\Big)_n$ converges to $\int_X d(x_0,x)^p \, d\mu(x)$.*

From this we can infer the first part of Theorem A.2 :

Theorem A.2.1. *Let (X,d) be a separable complete metric space and p a positive number. Then the w-$\mathcal{C}_{bp}(X)$ topology on $\mathcal{P}_p(X)$ is defined by the metric W_p.*

Proof. It stems from Proposition A.17 and the fact that the w-$\mathcal{C}_{bp}(X)$ topology is metrizable by Theorem A.1. □

Although $(\mathcal{P}_p(X), w\text{-}\mathcal{C}_{bp}(X))$ is topologically complete by Theorem A.1, the previous result does not necessarily ensure that $(\mathcal{P}_p(X), W_p)$ is complete.

We now prove here in several steps this result, which is the second part of Theorem A.2 (see also [3] or [93] for other proofs). First of all :

Proposition A.18. *Let (X,d) a separable complete metric space and $(\mu_n)_n$ a Cauchy sequence in $(\mathcal{P}_1(X), W_1)$. Then $(\mu_n)_n$ is uniformly tight.*

Proof. We adapt the classical proof of Ulam lemma (Proposition A.7).

ε being a given positive number, there exists N such that

$$W_1(\mu_n, \mu_N) \leq \varepsilon^2$$

for any $n \geq N$. In other words, for any n, there exists $j \leq N$ such that

$$W_1(\mu_n, \mu_j) \leq \varepsilon^2. \tag{A.3}$$

The finite family $(\mu_j)_{j \leq N}$ is uniformly tight by Ulam lemma, so there is a compact set K of X such that

$$\mu_j(K) \geq 1 - \varepsilon$$

for any $j \leq N$. Since K is compact, there exist p points $x_1, \dots, x_p$ in X such that

$$K \subset U = \bigcup_{k=1}^{p} B(x_k, \varepsilon)$$

where $B(x_k, \varepsilon) = \{x \in X; d(x_k, x) < \varepsilon\}$. In particular

$$\mu_j(U) \geq 1 - \varepsilon \tag{A.4}$$

for any $j \leq N$.

But by Kantorovich-Rubinstein formulation of W_1

$$W_1(\mu, \nu) = \sup \left\{ \left| \int_X f(x)\, d\mu(x) - \int_X f(x)\, d\nu(x) \right| ; [f]_{lip} \leq 1 \right\}$$

where the Lipschitz seminorm is defined by

$$[f]_{lip} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)},$$

we have in particular

$$\int_X \phi(x)\, d\mu_j(x) \; - \; \int_X \phi(x)\, d\mu_n(x) \; \leq \; \frac{1}{\varepsilon} W_1(\mu_j, \mu_n)$$

for any j and n, where ϕ is the $\dfrac{1}{\varepsilon}$-Lipschitz function defined on X by $\phi(x) = \left(1 - \frac{d(x,U)}{\varepsilon}\right)^+$.

On the other hand $\mathbf{1}_U \leq \phi \leq \mathbf{1}_{U^\varepsilon}$ where $U^\varepsilon = \{x; d(x, U) < \varepsilon\}$, so

$$\int_X \phi(x)\, d\mu_j(x) \geq \mu_j(U) \qquad \text{and} \qquad \int_X \phi(x)\, d\mu_n(x) \leq \mu_n(U^\varepsilon).$$

Consequently

$$\mu_n(U^\varepsilon) \; \geq \; \mu_j(U) - \frac{1}{\varepsilon} W_1(\mu_j, \mu_n) \tag{A.5}$$

for any j and n.

But by (A.3), for any n there exists $j \leq N$ such that $W_1(\mu_n, \mu_j) \leq \varepsilon^2$. Then by (A.4) and (A.5)

$$\mu_n(U^\varepsilon) \geq 1 - 2\,\varepsilon.$$

Thus for any $\varepsilon > 0$ we have found a set U^ε such that $\mu_n(U^\varepsilon) \geq 1 - 2\,\varepsilon$ for any n. But U^ε is not necessarily (relatively) compact, though bounded, so at this point we can not claim that $(\mu_n)_n$ is uniformly tight. Now we are going to build a compact set $\overline{S}$ for which the same kind of property holds.

Up to now we have proven that for any $\varepsilon > 0$ there exist p points $x_1, \ldots, x_p$ such that

$$\mu_n\big(X \setminus \bigcup_{k=1}^{p} B(x_k, 2\varepsilon)\big) \;\leq\; 2\,\varepsilon$$

for any n since $U^\varepsilon \subset \bigcup_{k=1}^{p} B(x_k, 2\,\varepsilon)$.

For any integer m we then proceed as above with $\varepsilon\, 2^{-m-1}$ instead of ε. More precisely for any m we get a finite number $p(m)$ of points $x_1^m, \ldots, x_{p(m)}^m$ of X such that

$$\mu_n\big(X \setminus \bigcup_{k=1}^{p(m)} B\left(x_k^m, \varepsilon\, 2^{-m}\right)\big) \;\leq\; \varepsilon\, 2^{-m}$$

for any n.

Then we note that the set

$$S \;=\; \bigcap_{m=1}^{+\infty} \bigcup_{k=1}^{p(m)} B\left(x_k^m, \varepsilon\, 2^{-m}\right)$$

is totally bounded since for any ρ, choosing m such that $\varepsilon\, 2^{-m} \leq \rho$, we see that S is covered by the $p(m)$ balls $B(x_k^m, \varepsilon\, 2^{-m})$ with radius $\varepsilon\, 2^{-m} \leq \rho$. From this it follows that the closure $\overline{S}$ of S is compact since X is complete.

Moreover for any n

$$\mu_n(X \setminus S) \leq \sum_{m=1}^{+\infty} \mu_n\big(X \setminus \bigcup_{k=1}^{p(m)} B\left(x_k^m, \varepsilon\, 2^{-m}\right)\big) \leq \sum_{m=1}^{+\infty} \varepsilon\, 2^{-m} \;=\; \varepsilon.$$

Thus $\overline{S}$ is compact and satisfies $\mu_n(X \setminus \overline{S}) \leq \varepsilon$ for any n. In other words the sequence $(\mu_n)_n$ is uniformly tight. $\square$

Corollary A.19. *Let (X, d) be a separable complete metric space, $p \geq 1$ and $(\mu_n)_n$ be a Cauchy sequence in $(\mathcal{P}_p(X), W_p)$. Then there exists a subsequence of $(\mu_n)_n$ converging to a probability measure μ in w-$\mathcal{C}_b(X)$ sense.*

Proof. $(\mu_n)_n$ is Cauchy in $(\mathcal{P}_1(X), W_1)$ since $W_1 \leq W_p$, so is uniformly tight by Proposition A.18, and thus relatively sequentially compact in $\mathcal{P}(X)$ for the w-$\mathcal{C}_b(X)$ topology by Prokhorov theorem (Proposition A.6). □

Proposition A.20. *Let (X, d) be a separable complete metric space, $p \geq 1$ and $(\mu_{n'})_{n'}$ a bounded sequence in $(\mathcal{P}_p(X), W_p)$, converging to a probability measure μ for the w-$\mathcal{C}_b(X)$ topology. Then μ also belongs to $\mathcal{P}_p(X)$.*

Proof. Let x_0 be a given point in X. Then

$$\int_X d(x_0, x)^p \, d\mu_{n'}(x) \, \leq \, 2^{p-1} \left(\int_X d(x_0, x)^p \, d\mu_0(x) + W_p(\mu_{n'}, \mu_0) \right) \, \leq \, C$$

where C is a constant independent of n' since the sequence $\left(W_p(\mu_{n'}, \mu_0)\right)_{n'}$ is bounded.

On the other hand $(\mu_{n'})_{n'}$ converges to μ in w-$\mathcal{C}_b(X)$ sense, so by Proposition A.16 we get

$$\int_X d(x_0, x)^p \, d\mu(x) \, \leq \, \liminf_{n' \to +\infty} \int_X d(x_0, x)^p \, d\mu_{n'}(x) \, \leq C,$$

which ensures that indeed $\mu \in \mathcal{P}_p(X)$. □

Now we give the proof of the second part of Theorem A.2 :

Theorem A.2.2. *If (X,d) is a separable complete metric space and p a positive number, then the metric space $(\mathcal{P}_p(X), W_p)$ is complete.*

Proof. 1. We first consider the case $p \geq 1$.

Let indeed $(\mu_n)_n$ be a Cauchy sequence in $(\mathcal{P}_p(X), W_p)$. Then by Corollary A.19 there exists a subsequence $(\mu_{n'})_{n'}$ converging to a probability measure μ in w-$\mathcal{C}_b(X)$ sense, with μ in $\mathcal{P}_p(X)$ by Proposition A.20.

We now prove that $W_p(\mu, \mu_{n'})$ tends to 0 as n' goes to infinity. The sequence $(\mu_{n'})_{n'}$ converges to μ in w-$\mathcal{C}_b(X)$ sense, so by Proposition A.16

$$W_p(\mu, \mu_{m'}) \leq \liminf_{n' \to +\infty} W_p(\mu_{n'}, \mu_{m'}). \tag{A.6}$$

for any given m' in the sequence (n').

But $(\mu_{n'})_{n'}$ is Cauchy for the distance W_p, so for any $\varepsilon > 0$, and n', m' large enough

$$W_p(\mu_{n'}, \mu_{m'}) \leq \varepsilon. \tag{A.7}$$

From (A.6) and (A.7) it finally follows that

$$W_p(\mu, \mu_{m'}) \leq \varepsilon$$

for m' large enough, which means that indeed $W_p(\mu, \mu_{n'})$ tends to 0 as n' goes to $+\infty$.

Finally $W_p(\mu_n, \mu)$ tends to 0 as n goes to infinity since the whole sequence $(\mu_n)_n$ is Cauchy in $(\mathcal{P}_p(X), W_p)$.

2. Then we consider the case $0 < p < 1$.

d^p is a distance on X which defines the same topology as d, so that, with obvious notations, we have $\mathcal{P}_p(X, d) = \mathcal{P}_1(X, d^p)$ and $\mathcal{C}_{bp}(X, d) = \mathcal{C}_{b1}(X, d^p)$.

Moreover (X, d^p) is complete if (X, d) is complete and $W_p(X, d) = W_1(X, d^p)$.

Thus, given a weight $p \in]0, 1[$ and a metric d on X, the results associated with the weight p and the metric d stem from the results associated with the weight 1 and the metric d^p proven in the previous step. □

Bibliographie

[1] L. Ambrosio : Lecture notes on optimal transport problems. *In Mathematical aspects of evolving surfaces (Funchal, 2000)*, vol. 1812 de *Lecture Notes in Math.* Springer, Berlin, 2003.

[2] L. Ambrosio, N. Gigli et G. Savaré : Gradient flows with metric and differentiable structures, and applications to the Wasserstein space. *Atti Acad. Naz. Lincei Cl. Sci. Rend. Lincei (9)*, 15, 3-4, p. 327–343, 2004. Disponible à l'adresse `http://cvgmt.sns.it/people/ambrosio`.

[3] L. Ambrosio, N. Gigli et G. Savaré : *Gradient flows in metric spaces and in the spaces of probability measures.* Birkhäuser, Basel, 2005.

[4] L. Ambrosio et A. Pratelli : Existence and stability results in the L^1 theory of optimal transportation. *In Optimal transportation and applications (Martina Franca, 2001)*, vol. 1813 de *Lecture Notes in Math.* Springer, Berlin, 2003.

[5] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto et G. Scheffer : *Sur les inégalités de Sobolev logarithmiques*, vol. 10 de *Panoramas et Synthèses.* Société Mathématique de France, Paris, 2000.

[6] A. Araujo et E. Giné : *The central limit theorem for real and Banach valued random variables.* John Wiley & Sons, New York, 1980.

[7] D. Bakry : On Sobolev and logarithmic Sobolev inequalities for Markov semigroups. *In New trends in stochastic analysis (Charingworth, 1994)*, Taniguchi Symposium. World Scientific, 1997.

[8] D. Bakry et M. Emery : Diffusions hypercontractives. *In Séminaire de probabilités XIX*, vol. 1123 de *Lecture Notes in Math.* Springer, Berlin, 1985.

[9] S. Benachour, B. Roynette, D. Talay et P. Vallois : Nonlinear self-stabilizing processes. I. Existence, invariant probability, propagation of chaos. *Stochastic Process. Appl.*, 75, 2, p. 173–201, 1998.

[10] S. Benachour, B. Roynette et P. Vallois : Nonlinear self-stabilizing processes. II. Convergence to invariant probability. *Stochastic Process. Appl.*, 75, 2, p. 203–224, 1998.

[11] D. Benedetto, E. Caglioti, J. A. Carrillo et M. Pulvirenti : A non-Maxwellian steady distribution for one-dimensional granular media. *J. Statist. Phys.*, 91, 5-6, p. 979–990, 1998.

[12] D. Benedetto, E. Caglioti et M. Pulvirenti : A kinetic equation for granular media. *RAIRO Modél. Math. Anal. Numér.*, 31, 5, p. 615–641, 1997.

[13] P. Billingsley : *Convergence of probability measures.* John Wiley & Sons, New York, 1999.

[14] G. Blower : The Gaussian isoperimetric inequality and transportation. *Positivity*, 7, p. 203–224, 2003.

[15] G. Blower et F. Bolley : Concentration of measure on product spaces with applications to Markov processes. Prépublication, 2005.

[16] S. G. Bobkov, I. Gentil et M. Ledoux : Hypercontractivity of Hamilton-Jacobi equations. *J. Math. Pures Appl.*, 80, 7, p. 669–696, 2001.

[17] S. G. Bobkov et F. Götze : Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163, p. 1–28, 1999.

[18] S. G. Bobkov et M. Ledoux : From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.*, 10, 5, p. 1028–1052, 2000.

[19] F. Bolley : Autour de l'équation de Boltzmann. Mémoire de magistère, Ecole normale supérieure, Paris. Disponible à l'adresse `http://www.umpa.ens-lyon.fr/~fbolley`, 2001.

[20] F. Bolley : Quantitative concentration inequalities on sample path space for mean field interaction. Prépublication, 2005.

[21] F. Bolley, Y. Brenier et G. Loeper : Contractive metrics for scalar conservation laws. *J. Hyperbolic Diff. Eq.*, 2, 1, p. 91–107, 2005.

[22] F. Bolley, A. Guillin et C. Villani : Quantitative concentration inequalities for empirical measures on non-compact spaces. A paraître dans *Prob. Theo. Rel. Fields*, 2005.

[23] F. Bolley et C. Villani : Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Ann. Fac. Sci. Toulouse*, 6, 14, 3, p. 331–352, 2005.

[24] N. Bourbaki : *Eléments de mathématique, Intégration, Chapitre IX.* Hermann, Paris, 1969.

[25] W. Braun et K. Hepp : The Vlasov dynamics and its fluctuations in the $1/n$ limit of interacting classical particles. *Comm. Math. Phys.*, 56, 2, p. 101–113, 1977.

[26] Y. Brenier : Résolution d'équations d'évolution quasilinéaires en dimension N d'espace à l'aide d'équations linéaires en dimension $N+1$. *J. Diff. Eq.*, 50, 3, p. 375–390, 1983.

[27] Y. Brenier : Averaged multivalued solutions and scalar conservation laws. *SIAM J. Numer. Anal.*, 21, 6, p. 1013–1037, 1984.

[28] Y. Brenier : Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44, 4, p. 375–417, 1991.

[29] Y. Brenier : Le système de Born-Infeld élargi : des ondes aux particules et aux cordes. Séminaire X-EDP, Ecole Polytechnique, 2005.

[30] Y. Brenier et E. Grenier : Sticky particles and scalar conservation laws. *SIAM J. Numer. Anal.*, 35, 6, p. 2317–2328, 1998.

[31] L. A. Caffarelli : Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.*, 214, 3, p. 537–563, 2000 and 225, 2, p. 449–450 (2002).

[32] E. Caglioti et F. Rousset : Long time estimates in the mean field limit. Prépublication, 2004.

[33] E. Carlen et W. Gangbo : Constrained steepest descent in the 2-Wasserstein metric. *Ann. of Math.*, 157, 3, p. 807–846, 2003.

[34] J. A. Carrillo, M. D. Francesco et G. Toscani : Strict contractivity of the 2-Wasserstein distance for the porous medium equation by mass-centering. Prépublication, 2005.

[35] J. A. Carrillo, M. P. Gualdani et G. Toscani : Finite speed of propagation on porous media by mass transportation methods. *C. R. Acad. Sci. Paris, Série I Math.*, 338, 10, p. 815–818, 2004.

[36] J. A. Carrillo, R. J. McCann et C. Villani : Kinetic equilibration rates for granular media and related equations : entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana*, 19, 3, p. 971–1018, 2003.

[37] J. A. Carrillo, R. J. McCann et C. Villani : Contractions in the 2-Wasserstein length space and thermalization of granular media. A paraître dans *Arch. Rat. Mech. Anal.*, 2005.

[38] J. A. Carrillo et G. Toscani : Wasserstein metric and large-time asymptotics of non-linear diffusion equations. *In New trends in mathematical physics (in honour of the Salvatore Rionero 70th birthday)*. World Scientific, Singapore, 2005.

[39] P. Cattiaux et A. Guillin : Talagrand's like quadratic transportation cost inequalities. Prépublication, 2005.

[40] D. Cordero-Erausquin : Some applications of mass transport to Gaussian-type inequalities. *Arch. Rat. Mech. Anal.*, 161, 3, p. 257–269, 2002.

[41] C. Dafermos : *Hyperbolic conservation laws in continuum physics*, vol. 325 de *Grundlehren Math. Wiss.* Springer, Heidelberg, 2000.

[42] C. Dellacherie et P. A. Meyer : *Probabilities and potential*, vol. 29 de *North-Holland Mathematics Studies.* North-Holland Publishing Co., Amsterdam, 1978.

[43] A. Dembo et O. Zeitouni : *Large Deviations Techniques And Applications.* Springer, New York, 1998.

[44] L. Desvillettes et C. Villani : On the spatially homogeneous Landau equation for hard potentials. I. Existence, uniqueness and smoothness. *Comm. PDE*, 25, 1-2, p. 179–259, 2000.

[45] M. DiFrancesco et P. A. Markowich : Entropy dissipation and Wasserstein metric methods for the viscous Burgers' equation : convergence to diffusive waves. Prépublication, 2004.

[46] H. Djellout, A. Guillin et L. Wu : Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.*, 32, 3B, p. 2702–2732, 2004.

[47] R. M. Dudley : *Real analysis and probability.* Cambridge University Press, Cambridge, 2002.

[48] J. Feng et T. G. Kurtz : *Large deviations for stochastic processes.* A paraître, 2005.

[49] X. Fernique : Régularité des trajectoires des fonctions aléatoires gaussiennes. *In Ecole d'Eté de Probabilités de Saint-Flour 1974*, vol. 480 de *Lecture Notes in Math.* Springer, Berlin, 1975.

[50] X. Fernique : Sur le théorème de Kantorovitch-Rubinstein dans les espaces polonais. *In Séminaire de Probabilités XV*, vol. 850 de *Lecture Notes in Math.* Springer, Berlin, 1981.

[51] H. Föllmer : Stock price fluctuation as a diffusion in a random environment. *In Mathematical Models in Finance. Eds S.D. Howison, F.P. Kelly and P. Wilmott.* Chapman and Hall, London, 1995.

[52] W. Gangbo et R. J. McCann : The geometry of optimal transportation. *Acta Math.*, 177, 2, p. 113–161, 1996.

[53] F. Gao : Moderate deviations and large deviations for kernel density estimators. *J. Theor. Prob.*, 16, p. 401–418, 2003.

[54] I. Gentil : Inégalités de Sobolev logarithmiques et de Poincaré pour la loi uniforme. Note disponible à l'adresse `http://www.ceremade.dauphine.fr/~gentil`, 2004.

[55] E. Giné et J. Zinn : Empirical processes indexed by Lipschitz functions. *Ann. Probab.*, 14, 4, p. 1329–1338, 1986.

[56] N. Gozlan : Principe conditionnel de Gibbs pour des contraintes fines approchées et inégalités de transport. Thèse de doctorat de l'Université de Paris 10 Nanterre, 2005.

[57] N. Gozlan et C. Léonard : A large deviation approach to some concentration cost inequalities. Prépublication, 2005.

[58] L. Gross : Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97, p. 1061-1083, 1975.

[59] N. Grunewald, F. Otto, M. G. Reznikoff et C. Villani : A two-scale proof of a Logarithmic Sobolev Inequality. Prépublication, 2005.

[60] M. Hauray et P. E. Jabin : N particles approximation of the Vlasov equations with singular potentials. A paraître dans *Arch. Rat. Mech. Anal.*, 2005.

[61] L. Hörmander : *Lectures on nonlinear hyperbolic differential equations*, vol. 26 de *Mathematics & Applications.* Springer, Berlin, 1997.

[62] R. Jordan, D. Kinderlehrer et F. Otto : The variationnal formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29, 1, p. 1–17, 1998.

[63] M. Kac : *Probability and related topics in physical sciences.* Interscience, London, 1959.

[64] D. Kinderlehrer et N. W. Walkington : Approximation of parabolic equations using the Wasserstein metric. *M2AN Math. Model. Numer. Anal.*, 33, p. 837–852, 1999.

[65] S. Kružkov : Generalized solutions of the Cauchy problem in the large for first order nonlinear equations. *Dokl. Akad. Nauk. SSSR*, 187, p. 29–32, 1969.

[66] S. R. Kulkarni et O. Zeitouni : A general classification rule for probability measures. *Ann. Stat.*, 23, 4, p. 1393–1407, 1995.

[67] M. Ledoux : Concentration of measure and logarithmic Sobolev inequalities. *In Séminaire de Probabilités XXXIII*, vol. 1709 de *Lecture Notes in Math.* Springer, Berlin, 1999.

[68] M. Ledoux : *The concentration of measure phenomenon*, vol. 89 de *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, 2001.

[69] M. Ledoux : Logarithmic Sobolev inequalities for unbounded spin systems revisited. *In Séminaire de Probabilités XXXV*, vol. 1755 de *Lecture Notes in Math.* Springer, Berlin, 2001.

[70] M. Ledoux et M. Talagrand : *Probability in Banach spaces.* Springer, Berlin, 1991.

[71] H. Li et G. Toscani : Long-time asymptotics of kinetic models of granular flows. *Arch. Rat. Mech. Anal.*, 172, 3, p. 407–428, 2004.

[72] G. G. Lorentz : *Approximation of functions.* Athena series, Selected topics in mathematics. Holt, Rinehart and Winston, New York, 1966.

[73] F. Malrieu : Inégalités de Sobolev logarithmiques pour des problèmes d'évolution non linéaires. Thèse de doctorat de l'Université Paul Sabatier, Toulouse, 2001.

[74] F. Malrieu : Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stochastic Process. Appl.*, 95, 1, p. 109–132, 2001.

[75] F. Malrieu : Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.*, 13, 2, p. 540–560, 2003.

[76] C. Marchioro et M. Pulvirenti : *Mathematical theory of incompressible nonviscous fluids.* Springer, New York, 1994.

[77] K. Marton : Bounding $\overline{d}$-distance by information divergence : a method to prove measure concentration. *Ann. Probab.*, 24, p. 857–866, 1996.

[78] K. Marton : A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6, p. 556–571, 1996.

[79] K. Marton : Measure concentration for Euclidean distance in the case of dependent random variables. *Ann. Probab.*, 32, p. 2526–2544, 2004.

[80] K. Marton : Logarithmic Sobolev inequality for weakly dependent random variables. Prépublication, 2005.

[81] P. Massart : Notes from a course given in Ecole d'Eté de Probabilités de Saint-Flour 2003. Disponible à l'adresse `http://www.math.u-psud.fr/~massart`, 2003.

[82] R. J. McCann : Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80, 2, p. 309–323, 1995.

[83] R. J. McCann : Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11, p. 589–608, 2001.

[84] S. Méléard : Asymptotic behaviour of some interacting particle systems ; McKean-Vlasov and Boltzmann models. *In Probabilistic models for nonlinear partial differential equations, Montecatini Terme, 1995*, vol. 1627 de *Lecture Notes in Math.* Springer, Berlin, 1996.

[85] H. Neunzert : An introduction to the nonlinear Boltzmann-Vlasov equation. *In Kinetic theories and the Boltzmann equation*, vol. 1048 de *Lecture Notes in Math.* Springer, Berlin, 1984.

[86] F. Otto : Dynamics of labyrinthine pattern formation in magnetic fluids : a mean-field theory. *Arch. Rat. Mech. Anal.*, 141, p. 63–103, 1998.

[87] F. Otto : The geometry of dissipative evolution equations : the porous medium equation. *Comm. PDE*, 26, 1-2, p. 101–174, 2001.

[88] F. Otto et C. Villani : Generalization of an inequality by Talagrand, and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173, p. 361–400, 2000.

[89] V. I. Paulsen : *Completely Bounded Maps and Dilations.* Longman Science and Technical, Harlow, 1986.

[90] V. V. Petrov : *Limit theorems of probability theory.* The Clarendon Press Oxford University Press, New York, 1995.

[91] R. Philipowski : Interacting diffusions approximating the porous medium equation and propagation of chaos. Prépublication, 2005.

[92] Y. Prohorov : Convergence of random processes and limit theorems in probability theory. *Th. Prob. and its Appl.*, 1, p. 157–214, 1956.

[93] S. T. Rachev : *Probability Metrics and the Stability of Stochastic Models.* John Wiley and Sons, Chichester, 1991.

[94] S. T. Rachev et L. Rüschendorf : *Mass transportation problems. Vol. I and II.* Springer, New York, 1998.

[95] M. Rosenblatt : *Markov Processes : Structure and Asymptotic Behavior.* Springer, Heidelberg, 1971.

[96] P.-M. Samson : Concentration of measure inequalities for Markov chains and ϕ-mixing processes. *Ann. Probab.*, 28, p. 416–461, 2000.

[97] S. Schochet : The weak vorticity formulation of the 2-D Euler equations and concentration-cancellation. *Comm. PDE*, 20, 5-6, p. 1077–1104, 1995.

[98] S. Schochet : The point-vortex method for periodic weak solutions of the 2-D Euler equations. *Comm. Pure Appl. Math.*, 49, 9, p. 911–965, 1996.

[99] L. Schwartz : *Topologie générale et analyse fonctionnelle.* Hermann, Paris, 1970.

[100] L. Schwartz : *Radon measures on arbitrary topological spaces and cylindrical measures.* Published for the Tata Institute of Fundamental Research, Bombay by Oxford University Press, London, 1973.

[101] D. Serre : *Systems of conservation laws I.* Cambridge University Press, Cambridge, 1999.

[102] D. Serre : L^1-stability of nonlinear waves in scalar conservation laws. *In Hanbook of differential equations. Evolutionary equations, volume 1. Eds C. M. Dafermos and E. Feireisl.* Elsevier, 2004.

[103] M. Sion : On general minimax theorems. *Pac. J. Math.*, 8, p. 171–176, 1958.

[104] H. Spohn : *Large scale dynamics of interacting particles.* Texts and Monographs in Physics. Springer, Berlin, 1991.

[105] A.-S. Sznitman : Topics in propagation of chaos. *In Ecole d'Eté de Probabilités de Saint-Flour 1989*, vol. 1464 de *Lecture Notes in Math.* Springer, Berlin, 1991.

[106] M. Talagrand : Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6, p. 587–600, 1996.

[107] D. Talay : Probabilistic numerical methods for partial differential equations : elements of analysis. *In Probabilistic models for nonlinear partial differential equations, Montecatini Terme, 1995*, vol. 1627 de *Lecture Notes in Math.* Springer, Berlin, 1996.

[108] H. Tanaka : Probabilistic treatment of the Boltzmann equation of Maxwellian molecules. *Z. Wahr. Verw. Gebiete*, 46, 1, p. 67–105, 1978-79.

[109] H. Tanaka : Limit theorems for certain diffusion processes with interaction. *In Stochastic analysis, Kataka/Tokyo, 1982.* North Holland, Amsterdam, 1984.

[110] A. van der Vaart et J. Wellner : *Weak convergence and empirical processes.* Springer, Berlin, 1995.

[111] C. Villani : *Topics in optimal transportation*, vol. 58 de *Grad. Stud. Math.* American Mathematical Society, Providence, 2003.

[112] F.-Y. Wang : Logarithmic Sobolev inequalities on noncompact Riemannian manifolds. *Prob. Theo. Rel. Fields*, 108, p. 417–424, 1997.

[113] F.-Y. Wang : Probability distance inequalities on Riemannian manifolds and path spaces. *J. Funct. Anal.*, 206, 1, p. 167–190, 2004.